

GenEmb1:*

1: gb.pa:*

2: gb.hng:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vt:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.pl:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vt:*

30: em.hng.hum:*

31: em.hng.inv:*

32: em.hng.other:*

33: em.hng.pls:*

34: em.hng.pln:*

35: em.hng.rod:*

36: em.hng.mam:*

37: em.hng.yrtr:*

38: em.sy:*

39: em.hgo.hum:*

40: em.hgo.mus:*

41: em.hgo.other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match Length	DB	ID	Description
1	860.4	58.6	1574	9	BC012868 Homo sapi
2	829	95.0	1628	6	AX014903 Sequence
3	827	94.7	1558	9	BC001668 Homo sapi
4	825.4	94.5	1424	9	BC001855 Homo sapi
5	825.4	94.5	1457	9	BC001951 Homo sapi
6	825.4	94.5	1509	9	BC007705 Homo sapi
7	825.2	94.5	1544	9	AX010783 Sequence
8	825.2	94.5	1544	9	AB009519 Homo sapi
9	809.4	92.7	1512	9	AB006534 Homo sapi
10	805.6	92.3	1554	6	AX302243 Sequence
11	805.6	92.3	1544	9	AF027205 Homo sapi
12	621	71.1	16960	9	AY050668 Homo sapi
13	621	71.1	16960	9	AC001879 Homo sapi
14	595.2	68.2	64620	2	AC063794 Homo sapi
15	394	45.1	759	6	AF069031 Sequence
16	394	45.1	759	6	EL1900 Human CDNA
17	394	45.1	759	6	194991 Sequence (
18	337	38.6	349	9	HX5897 Homo sapiens
19	316.4	36.2	320	6	AX302973 Sequence
20	312.2	35.8	331	6	AX337128 Sequence
21	312.2	35.8	331	6	AX409401 Sequence
22	273	31.3	285	6	AX260381 Sequence
23	262	30.0	286	6	AX261052 Sequence
24	255.2	29.2	1237	10	BC003431 Mus muscu
25	253.8	29.1	1287	6	AX261672 Sequence
26	252	28.9	1174	10	BC026419 Mus muscu
27	252	28.9	1237	10	AF099019 Mus muscu
28	252	28.9	1310	10	AF099020 Mus muscu
29	252	28.9	1408	10	AF099016 Mus muscu
30	238.4	27.3	241	6	AX303056 Sequence
31	226.4	25.9	287	6	AX261607 Sequence
32	200	22.9	200	6	AX379141 Sequence
33	186.2	21.3	336	6	AX093200 Sequence
34	167.4	19.2	207	6	AX366602 Sequence
35	167	19.1	207	6	AX366589 Sequence
36	167	19.1	224	6	AX261692 Sequence
37	167	19.1	234	6	AX261412 Sequence
38	167	19.1	234	6	AX261478 Sequence
39	167	19.1	234	6	AX261740 Sequence
40	167	19.1	234	6	AX379315 Sequence
41	167	19.1	371	6	AX366598 Sequence
42	165.4	18.9	234	6	AX261539 Sequence
43	165.4	18.9	276	6	AX366401 Sequence
44	150.6	17.3	302	6	AX261340 Sequence
45	146.8	16.8	199	6	AX302805 Sequence

ALIGNMENTS

RESULT 1	
BC012868	
LOCUS	1574 bp mRNA linear PRI 22-Aug-2001
DEFINITION	Homo sapiens, serine protease inhibitor, Kunitz type, 2, clone
ACCESSION	MGC:9154 IMAGE:3857277, mRNA, complete cds.
VERSION	BC012868
KEYWORDS	BC012868.1 GI:15377533
SOURCE	MGC.
ORGANISM	Homo sapiens.
	Homo sapiens
REFERENCE	Bukacinska; Melazzo; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 1574)
	Strausberg,R.
	Direct Submission

JOURNAL

Submitted (20-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcgabs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HQSC
 Web site: <http://www.hqsc.bcm.tmc.edu/cdna/>
 Contact: villalobscbm.tmc.edu
 Villalon, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
 Series: IPAK Plate: 21 Row: b Column: 23
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 2598967.

FEATURES

source

1..1574
 /organism="Homo sapiens"
 /db_xref="locusid:10653"
 /db_xref="taxon:9606"
 /clone="MGC:9154 IMAGE:3857277"
 /tissue_type="Ovary, adenocarcinoma"
 /clone_id="NIH_MGC_66"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 335..1093
 /codon_start=1
 /product="serine protease inhibitor, Kunitz type, 2"
 /protein_id="AAH12868.1"
 /db_xref="GI:15277534"
 /translation="MAOLGLRRSAPFLALISLLISGLVAADRESIHDFCLVSKYVGRCAKSPRMVYVTDGSQLFVGGCGGNSNNITKEGKCKKATVENATGLPATS RNAADSSVPSAPRRQDSBDSHSDMNYEYCTANAYVGPGRASFPKMYFPIERRSCNN FYVGGCKGNKSYRSEACMLRCFQSQENPELPYVVLGLFVWLLFLGASNV YLIRARNRNOERALRTYVSSGDDKOLVXNYLV"

CDS

BASE COUNT 318 a 409 c 485 g 362 t
 ORIGIN

Query Match 98.6%; Score 860.4; DB 9; Length 1574;
 Best Local Similarity 99.8%; Pred. No. 4,4e-195;
 Matches 872; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 CTCAGGATATTTCACTATGAAGATCTGACGCGCAAGCACTGAGGCTTG 60
 DB 700 CTCAGGATATTTCACTATGAAGATCTGACGCGCAAGCACTGAGGCTTG 759
 QY 61 CCGTGATCTCTCCACGCTGCTACTTGAAGTGAAGAGAACTCTGCAATACTTCAT 120
 DB 760 CCGTGATCTCTCCACGCTGCTACTTGAAGTGAAGAGAACTCTGCAATACTTCAT 819
 QY 121 CTATGAGGCTGCGCGGCAATAGAACAGCTACCGCTCTGAGAGGCTTGATGCTCG 180
 DB 820 CTATGAGGCTGCGCGGCAATAGAACAGCTACCGCTCTGAGAGGCTTGATGCTCG 879
 QY 181 CTGCTTCCGCGACAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 DB 880 CTGCTTCCGCGACAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 939
 QY 241 GGGGCTGTGTGATGTGATCTCTTCTGAGAGCTCCAGAGTCTGATCCG 300
 DB 940 GGGGCTGTGTGATGTGATCTCTTCTGAGAGCTCCAGAGTCTGATCCG 999
 QY 301 GGTGCAAGAGAACCAAGAGCTGCTCTGCAAGCTCTGAGGCTCGGATGACAA 360

DB 1000 GGTGCAAGAGAACCAAGAGCTGCTCTGCAAGCTCTGAGGCTCGGATGACAA 1059
 QY 361 GGAGCACTGTGTGAAGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
 DB 1060 GGAGCACTGTGTGAAGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1119
 QY 420 AAGGAGGAGAGATATGTGTGAGCTTTTAAATAGAGGATGAGTCTGATTTGAGT 479
 DB 1120 AAGGAGGAGAGATATGTGTGAGCTTTTAAATAGAGGATGAGTCTGATTTGAGT 1179
 QY 480 GATCATTAAGGCTGAGTCTGTTCTCTGAGAGTAGAGAGCTGCTTCTGCTGCTGCA 539
 DB 1180 GATCATTAAGGCTGAGTCTGTTCTCTGAGAGTAGAGAGCTGCTTCTGCTGCTGCA 1239
 QY 540 GGGATGGGTTTCTGTTGAAATCTCTAGAGGCTCTCTGCTGCTGCTGCTGCTG 599
 DB 1240 GGGATGGGTTTCTGTTGAAATCTCTAGAGGCTCTCTGCTGCTGCTGCTGCTG 1299
 QY 600 CAGACACCCGAGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
 DB 1300 CAGACACCCGAGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1359
 QY 660 GCTTATGTTGAATTCATGCTCTTCTCTCATCACAGAGTATGTTGATGCTTCT 719
 DB 1360 GCTTATGTTGAATTCATGCTCTTCTCTCATCACAGAGTATGTTGATGCTTCT 1419
 QY 720 TTTGTTGCTGATTTATGTTTATTAAGTATAAACAAAGTTTATTTATGATTTCTG 779
 DB 1420 TTTGTTGCTGATTTATGTTTATTAAGTATAAACAAAGTTTATTTATGATTTCTG 1479
 QY 780 AAGAGGAACTAAATGTCAGTTTAAATAAGAGGCTTCCCTTATGATTAAT 839
 DB 1480 AAGAGGAACTAAATGTCAGTTTAAATAAGAGGCTTCCCTTATGATTAAT 1539
 QY 840 TTCAGATGCTTCAAAAAAAAAAAAAAAAAAAAA 873
 DB 1540 TTCAGATGCTTCAAAAAAAAAAAAAAAAAAAAA 1573

RESULT 2
 LOCUS AX014903 1628 bp DNA linear PAT 07-SEP-2000
 DEFINITION Sequence 102 from Patent WO953040.
 ACCESSION AX014903
 VERSION AX014903.1 GI:10041170
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 1628)
 Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
 Pilarsky, C.
 TITLE Human nucleic acid sequences from ovarian tumour tissue
 JOURNAL Patent: WO 953040-A-102.21-OCT-1999;
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
 BERND (DE); ROSENTHAL ANDRE (DE); METZGEN GERT FUER GENOMFORSCHUN
 (DE); PILARSKY CHRISTIAN (DE)
 FEATURES
 source
 1..1628
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 329 a 433 c 498 g 368 t
 ORIGIN

Query Match 95.0%; Score 829; DB 6; Length 1628;
 Best Local Similarity 98.2%; Pred. No. 1.4e-187;
 Matches 849; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 CTCAGGATATTTCACTATGAAGATCTGACGCGCAAGCACTGAGGCTTG 60
 DB 764 CTCAGGATATTTCACTATGAAGATCTGACGCGCAAGCACTGAGGCTTG 823

```

QY      61  CCGTGCATCCCTCCACGCGGTACTTTGAGCTGAGAGGAACCTCTGCATTAATCTCAT 120
DB      824  CCGTGACATCCCTCCACGCGGTACTTTGAGCTGAGAGGAACCTCTGCATTAATCTCAT 883
QY      121  CTATGAGAGGCTCCCGGGGCAATTAAGACAGCTACCGCTCTGAGAGAGGCTCGATCTCCG 180
DB      884  CTATGAGAGGCTCCCGGGGCAATTAAGACAGCTACCGCTCTGAGAGAGGCTCGATCTCCG 943
QY      181  CTGCTTCGCGCAGAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGGTGTCTGCGC 240
DB      944  CTGCTTCGCGCAGAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGGTGTCTGCGC 1003
QY      241  GGGGCTGTGTGATGTGTGATCTCTCTCTCTGAGAGCTCCATGATGTCTTACTGATCCG 300
DB      1004  GGGGCTGTGTGATGTGTGATCTCTCTCTCTGAGAGCTCCATGATGTCTTACTGATCCG 1063
QY      301  GGTGGCAACGAGAGAACAGAGAGGCGCCCTGGACCGCTGTGAGCTCCGAGATGACAA 360
DB      1064  GGTGGCAACGAGAGAACAGAGAGGCGCCCTGGACCGCTGTGAGCTCCGAGATGACAA 1123
QY      361  GGAGGAGCTGGTGAAGAACACATATGTCTGACCGCCCTGTGCGCAAGAGACT-GGG 419
DB      1124  GGAGGAGCTGGTGAAGAACACATATGTCTGACCGCCCTGTGCGCAAGAGACTGGGG 1183
QY      420  AAGGAGGGGAGACTATGTGTGAGCTTTTAAATAGAGAGATGATCGGATTTGAGT 479
DB      1184  AAGGAGGGGAGACTATGTGTGAGCTTTTAAATAGAGAGATGATCGGATTTGAGT 1243
QY      480  GATCATTTGAGGCTGAGGTCTGTCTCTGGAGAGTGAAGAGGCTCTCTGCTGCTGCA 539
DB      1244  GATCATTTGAGGCTGAGGTCTGTCTCTGGAGAGTGAAGAGGCTCTCTGCTGCTGCA 1303
QY      540  GGGATGGGTTTGGCTTTGAAAATCCTCTAGAGAGCTCCTCCTCGATAGCCTCGAGCTGG 599
DB      1304  GGGATGGGTTTGGCTTTGAAAATCCTCTAGAGAGCTCCTCCTCGATAGCCTCGAGCTGG 1353
QY      600  CAGCAGCCCGAGTTTCTCTGCTGATGCAATTTCTTCTCCAGGTAGAGTTTCTTT 659
DB      1364  CAGCAGCCCGAGTTTCTCTGCTGATGCAATTTCTTCTCCAGGTAGAGTTTCTTT 1423
QY      660  GCTTATGTGAATTCATTCGCTCTCTTTCTCATCAGAGATGATGTGATCTGTTCT 719
DB      1424  GCTTATGTGAATTCATTCGCTCTCTTTCTCATCAGAGATGATGTGATCTGTTCT 1483
QY      720  TTGTTGTTGCTGATTTATGTTTTTAAAGTAAACAAAGTTTTTATAGCATCTCG 779
DB      1484  TTGTTGTTGCTGATTTATGTTTTTAAAGTAAACAAAGTTTTTATAGCATCTCG 1543
QY      780  AAAGAGGAAAGTAAATGTACAAGTTTAATAAAAAGGGGCTTCCCTTTAGCAATTAAT 839
DB      1544  AAAGAGGAAAGTAAATGTACAAGTTTAATAAAAAGGGGCTTCCCTTTAGCAATTAAT 1603
QY      840  TTCAGCATGTGCTTTCAAAAAAAA 864
DB      1604  AAAAAAAAAAAAAAAAAAAAAA 1628

RESULT 3
BC001668      1558 bp      mRNA      linear      PRI 12-JUL-2001
LOCUS      Homo sapiens, serine protease inhibitor, Kunitz type, 2, clone
DEFINITION      MGC:2021 IMAGE:2959462, mRNA, complete cds.
ACCESSION      BC001668
VERSION      BC001668.1 GI:12804514
KEYWORDS      MGC.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 1558)
AUTHORS      Strausberg R.
TITLE      Direct Submission
JOURNAL      Submitted (16-JAN-2001) National Institutes of Health, Mammalian

```

```

REMARK      Gene Collection (MGC), Cancer Genomics Office, National Cancer
COMMENT      Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
              Contact: MGC help desk
              Email: cgabs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
              DNA Sequencing by: Institute for Systems Biology
              http://www.systemsbio.org
              contact: amadan@systemsbiology.org
              Amp Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
              Greene, Mark Ketterman and Anuradha Madan

FEATURES
source      1..1558
              /organism="Homo sapiens"
              /db_xref="locusID:10653"
              /db_xref="taxon:9606"
              /clone="MGC:2021 IMAGE:2959462"
              /tissue_type="Colon, adenocarcinoma"
              /clone_lib="NIH_MGC_15"
              /lab_host="DH10B-R"
              /note="Vector: pOTB7"
              341..1099
              /codon_start=1
              /product="serine protease inhibitor, Kunitz type, 2"
              /protein_id="AAH01668.1"
              /db_xref="GI:12804515"
              /translation="MAQLCGLRSGRAFLALGSLISGLVLADEBRSIHDFCLVSKYV
              GRBASMPWNYNTDSSCOLFYGGDDGNSNNYLFKECLKKATVBNATGDLATS
              RNADSSVPSARRODSBDHSDFNFBCTNATVGPGRASPPRYTFVENSQNN
              FYVGGGRKNSYRSEACMLRCRQGENPPLPLGSKVVLALGFMVLIIFGASNV
              YLIRVARNRQERALTWSSGDDKEQLVKNTVYL"

BASE COUNT      314 a      401 c      484 g      359 t

ORIGIN
Query Match      94.7%; Score 927, DB 9; Length 1558;
Best Local Similarity 99.9%; Pred. No. 4.3e-197;
Matches 838; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY      1  CTCGACCGATATGTTCAACTATGAAGATCTGCACGCCCAAGCAGTCACTGGGCTTG 60
DB      706  CTCGACCGATATGTTCAACTATGAAGATCTGCACGCCCAAGCAGTCACTGGGCTTG 765
QY      61  CCGTGATCTTCCCAAGCTGTGTACTTTGAGCGTGAAGAGAACTCTGCATTAATTCAT 120
DB      766  CCGTGATCTTCCCAAGCTGTGTACTTTGAGCGTGAAGAGAACTCTGCATTAATTCAT 825
QY      121  CTATGAGAGGCTCCCGGGCAATTAAGACAGCTACCGCTCTGAGAGGCTCGATCTCCG 180
DB      826  CTATGAGAGGCTCCCGGGCAATTAAGACAGCTACCGCTCTGAGAGGCTCGATCTCCG 885
QY      181  CTGCTTCGCGCAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGGTGTCTGAGC 240
DB      886  CTGCTTCGCGCAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGGTGTCTGAGC 945
QY      241  GGGGCTGTGTGATGTGTGATCTCTCTCTGAGAGCTCCATGATGTCTAATCTCG 300
DB      946  GGGGCTGTGTGATGTGTGATCTCTCTCTGAGAGCTCCATGATGTCTAATCTCG 1005
QY      301  GGTGGCAACGAGAGAACAGAGGCTGCTGCGACCGTGTGAGCTCCGAGATGACAA 360
DB      1006  GGTGGCAACGAGAGAACAGAGGCTGCTGCGACCGTGTGAGCTCCGAGATGACAA 1065
QY      361  GGAGGAGCTGGTGAAGAACATATGTCTGCTGCTGAGAGGCT-GGG 419

```

Db 1066 GGAAGAGCTGGTGAAGACATATCTCTGTGACCGCCCTGTGCCAAGAGACTGGGG 1125

QY 420 AAGGAGGAGGAGACTATGTGTGAGCTTTTAAATAGAGGATTTGACTGGATTGAGT 479

Db 1126 AAGGAGGAGGAGACTATGTGTGAGCTTTTAAATAGAGGATTTGACTGGATTGAGT 1185

QY 480 GATCATTTAGGGCTGAGGCTCTCTCTGGAGGGTGAAGAGGCTGCTCTCTGCTGCA 539

Db 1186 GATCATTTAGGGCTGAGGCTCTCTCTGGAGGGTGAAGAGGCTGCTCTCTGCTGCA 1245

QY 540 GGGATGGGTTTGGCTTTGGAATCTCTAGAGGCTCCCTCGGATGAGCTGAGTGG 599

Db 1246 GGGATGGGTTTGGCTTTGGAATCTCTAGAGGCTCCCTCGGATGAGCTGAGTGG 1305

QY 600 CAGCAGCCCGAGTGTGTCTCTGCTGATGCAATTTCTTCCCGAGGATGAGTTTCTT 659

Db 1306 CAGCAGCCCGAGTGTGTCTCTGCTGATGCAATTTCTTCCCGAGGATGAGTTTCTT 1365

QY 660 GCTTATGTTGAATTCATTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719

Db 1366 GCTTATGTTGAATTCATTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1425

QY 720 TTGTTTGTCTGATTATGCTTTTAAATAGGATTTTAAATAGGATTTTAAATAGG 779

Db 1426 TTGTTTGTCTGATTATGCTTTTAAATAGGATTTTAAATAGGATTTTAAATAGG 1485

QY 780 AAGGAGGAGGAGTAAATGTAACAAGTTTAAATAGGAGGCTTCCCTTTAGAAATA 838

Db 1486 AAGGAGGAGGAGTAAATGTAACAAGTTTAAATAGGAGGCTTCCCTTTAGAAATA 1544

RESULT 4

BC011955 1424 bp mRNA linear PRI 02-AUG-2001

LOCUS

DEFINITION Homo sapiens, serine protease inhibitor, Kunitz type, 2, clone

ACCESSION MG:11712 IMAGE:3857238, mRNA, complete cds.

VERSION BC011955

KEYWORDS MG:

SOURCE Homo sapiens.

ORGANISM

REFERENCE 1 (bases 1 to 1424)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HOSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalob@bcm.tmc.edu

Villalob, D.K., Luna, R.A., Hale, S.M., Huily, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, C., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at: <http://image.llnl.gov>

Series: IRAC Plate: 21 Row: a Column: 22

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gt: 2598967.

FEATURES

1. 1424

Location/Qualifiers

source

/organism="Homo sapiens"

Query Match 94.5%; Score 825.4; DB 9; Length 1424;

Best Local Similarity 99.8%; Pred. No. 1e-186;

Matches 837; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

BASE COUNT 292 a 367 c 426 g 339 t

ORIGIN

/db_xref="LocusID:10653"

/db_xref="taxon:9606"

/clone="MG:11712 IMAGE:3857238"

/issue="Ovary, adenocarcinoma"

/clone_lib="NIH MGC_66"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

208.966

/codon_start=1

/product="serine protease inhibitor, Kunitz type, 2"

/protein_id="AAH195.1"

/db_xref="GI:15080403"

/translation="MAOLGGRSRAPFALLGSLLSGLVLAADRESHDFCLYKRV

GRASMPRMVNTVDSCQLFVYGCDGNSNNLTLECKKCAVTENATGLATS

RNAADSVPSAPFRSDSDHSBDMFNEYCTAAVATGCPRAFPWYDVERNSCN

FLVGCGRGNKSYSEACMLRCFRQENPLPLGSKVVLVLAFLFVNLILFLGASW

YLIRARRNRRLRTYSSGDKOLYKNTYV"

QY 1 CTCGAGGATATGTTCACTATGAAGATCTGACCCGCAACGACATGAGCTTGG 60

Db 573 CTCGAGGATATGTTCACTATGAAGATCTGACCCGCAACGACATGAGCTTGG 632

QY 61 CCGTGCATCTCTCCACGCTGTAATTTGAGCTGAGAGAACTCTGCAATTAATTGAT 120

Db 633 CCGTGCATCTCTCCACGCTGTAATTTGAGCTGAGAGAACTCTGCAATTAATTGAT 692

QY 121 CTATGAGGCTGCTGGGGCAATAGAACAGTACCGCTCTGAGAGGCTGATGCTCG 180

Db 693 CTATGAGGCTGCTGGGGCAATAGAACAGTACCGCTCTGAGAGGCTGATGCTCG 752

QY 181 CTGCTCCGACAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

Db 753 CTGCTCCGACAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 812

QY 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300

Db 813 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 872

QY 301 GGTGACAGAGAGAACCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

Db 873 GGTGACAGAGAGAACCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 932

QY 361 GGAGGAGCTGTTGAACACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419

Db 933 GGAGGAGCTGTTGAACACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 992

QY 420 AAGGAGGAGGAGACTATGTGTGAGCTTTTAAATAGAGGATTTGACTGGATTGAGT 479

Db 993 AAGGAGGAGGAGACTATGTGTGAGCTTTTAAATAGAGGATTTGACTGGATTGAGT 1052

QY 480 GATCATTTAGGGCTGAGGCTCTCTCTGGAGGGTGAAGAGGCTGCTCTCTGCTGCA 539

Db 1053 GATCATTTAGGGCTGAGGCTCTCTCTGGAGGGTGAAGAGGCTGCTCTCTGCTGCA 1112

QY 540 GGGATGGGTTTGGCTTTGGAATCTCTAGAGGCTCCCTCGGATGAGCTGAGTGG 599

Db 1113 GGGATGGGTTTGGCTTTGGAATCTCTAGAGGCTCCCTCGGATGAGCTGAGTGG 1172

QY 600 CAGCAGCCCGAGTGTGTCTCTGCTGATGCAATTTCTTCCCGAGGATGAGTTTCTT 659

Db 1173 CAGCAGCCCGAGTGTGTCTCTGCTGATGCAATTTCTTCCCGAGGATGAGTTTCTT 1232

QY 660 GCTTATGTTGAATTCATTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719

Db 1233 GCTTATGTTGAATTCATTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1292

QY 720 TTGTTTGTCTGATTATGCTTTTAAATAGGATTTTAAATAGGATTTTAAATAGG 779


```

Db 1293 TTGCTTCTGATTTATGTTTATTAAGTATTAACAAAGCTTTTATTAAGCATTCG 1352
Qy 780 AAAGAAGAAAGTAAATGACAAAGTTTAAATAAAGAGGAGCTTCCCTTTAGATATA 838
Db 1353 AAAGAAGAAAGTAAATGACAAAGTTTAAATAAAGAGGAGCTTCCCTTTAGATATA 1411

RESULT 5
BC011951 1457 bp mRNA linear PRI 02-AUG-2001
LOCUS Homo sapiens, serine protease inhibitor, Kunitz type, 2, clone
DEFINITION MGC:17171 IMAGE:3856970, mRNA, complete cds.
ACCESSION BC011951.1 GI:15080390
VERSION BC011951.1
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominda; Homo.
REFERENCE 1 (bases 1 to 1457)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
REMARK COMMENT
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Huijk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
Series: IRAX Plate: 21 Row: a Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 2598967.
FEATURES
Source
1..1457
/organism="Homo sapiens"
/db_xref="locusid:10653"
/db_xref="taxon:9606"
/clone="MGC:17171 IMAGE:3856970"
/tissue_type="Ovary, adenocarcinoma"
/clone_id="NIH_MGC_66"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
236..994
/codon_start=1
/product="serine protease inhibitor, Kunitz type, 2"
/protein_id="AAH1951.1"
/db_xref="GI:15080391"
/translation="MAOLGRLRSRAFLALGLSLGLAADRSHIDFLVSKV
GRCRASMRMYNVTGSCQLFYGGCGGNSNLTKEBCLKKCATVBNATGLATS
RNAASVSVAAPRODSESSSPMYEYCTANAYGPPRASFPPMYDVRNSCN
FTYGGCRNKSTISSEBACMLRCFRQDMPPLPGSKVYVLLAGLFVWLLIFLQASV
YLIRVARNRERARLTVSSGDKELVANYTVL"

BASE COUNT 304 a 374 c 439 g 340 t
ORIGIN
Query March 94 5%; Score 825.4; DB 9; Length 1457;
Best Local Similarity 99.8%; Pred. No. 1e-186;
Matches 937; Conservative 0; Mismatches 1; Indels 1;

```

```

Qy 1 CTCACGCAATGATGCTCAATATGAAGATATACGACCGGCAAGCAGTACCTGGGCTTGG 60
Db 601 CTCACGCAATGATGCTCAATATGAAGATATACGACCGGCAAGCAGTACCTGGGCTTGG 660
Qy 61 CCGTGATCTCTCCCAAGCTGGTACTTTGACGTGGAGAGAACTCTGCATTAATCTTCAT 120
Db 661 CCGTGATCTCTCCCAAGCTGGTACTTTGACGTGGAGAGAACTCTGCATTAATCTTCAT 720
Qy 121 CTATGAGAGCTGCGGAGCAATAAAGACACTACCGCTCTGAGAGAGCTCGATGCTCCG 180
Db 721 CTATGAGAGCTGCGGAGCAATAAAGACACTACCGCTCTGAGAGAGCTCGATGCTCCG 780
Qy 181 CTGCTTCCGCGACGAGAGAAATCTCCCTGCGCCCTTGAGCTCAAGGTGGTGTTCGTGC 240
Db 781 CTGCTTCCGCGACGAGAGAAATCTCCCTGCGCCCTTGAGCTCAAGGTGGTGTTCGTGC 840
Qy 241 GGGGCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Db 841 GGGGCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Qy 301 GATGACAGAGAGAAACAGAGAGCTGCGGCAAGCTCTGAGAGCTCGGAGATGACAA 360
Db 901 GATGACAGAGAGAAACAGAGAGCTGCGGCAAGCTCTGAGAGCTCGGAGATGACAA 960
Qy 361 GAGAGAGCTGAGAGAGAAACATATGCTCTGAGAGAGCTCTGAGAGAGCT-GGG 419
Db 961 GAGAGAGCTGAGAGAGAAACATATGCTCTGAGAGAGCTCTGAGAGAGCT-GGG 1020
Qy 420 AAGGAGAGGAGAGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 479
Db 1021 AAGGAGAGGAGAGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
Qy 480 GATCATTAAGAGCTGAGAGCTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
Db 1081 GATCATTAAGAGCTGAGAGCTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 540 GGGATGGGTTTCTTGGAAAATCTCTGAGAGAGCTCTCTGAGAGAGCTCTGAGAGAG 599
Db 1141 GGGATGGGTTTCTTGGAAAATCTCTGAGAGAGCTCTCTGAGAGAGCTCTGAGAGAG 1200
Qy 600 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
Db 1201 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 660 GCTTATGTAATTCATGCTGCTTTTCTCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Db 1261 GCTTATGTAATTCATGCTGCTTTTCTCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 720 TTGCTTCTGATTTATGATGATTTTATTAAGTATTAACAAAGCTTTTATTAAGCATTCG 779
Db 1321 TTGCTTCTGATTTATGATGATTTTATTAAGTATTAACAAAGCTTTTATTAAGCATTCG 1380
Qy 780 AAAGAAGAAAGTAAATGACAAAGTTTAAATAAAGAGGAGCTTCCCTTTAGATATA 838
Db 1381 AAAGAAGAAAGTAAATGACAAAGTTTAAATAAAGAGGAGCTTCCCTTTAGATATA 1439

RESULT 6
BC007705 1509 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, serine protease inhibitor, Kunitz type, 2, clone
DEFINITION MGC:11327 IMAGE:3952243, mRNA, complete cds.
ACCESSION BC007705
VERSION BC007705.1 GI:14043429
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominda; Homo.
REFERENCE 1 (bases 1 to 1509)
AUTHORS Strausberg, R.
TITLE Direct Submission

```

JOURNAL

Submitted (11-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-x@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nsl.nih.gov
Shcherbko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brickley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantip, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>
Series: RAL Plate: 15 Row: n Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

Location/Qualifiers
1..1509
/organism="Homo sapiens"
/db_xref="LOCUSID:10653"
/db_xref="taxon:9606"
/clone="MGC:11327 IMAGE:3952243"
/tissue_type="Ovary, adenocarcinoma"
/clone_id="NIH MGC 9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
289..1047
/codon_start=1
/product="serine protease inhibitor, Kunitz type, 2"
/protein_id="AAH07705.1"
/db_xref="GI:14043430"
/translation="MAQGLRSPRAFLALGSLLSGLVALADRSIHDFVSKVY
GRCASMRWYNTDSSCOLFYGGCGSNVYLTBECKKCAATNATGADTATS
KNAUDSVSPARQDSSEDSMEYETCTANNAVYGPGRASPAPVADVERNCNN
FIYGGKGNKNSYSEACMLRCFRQENPPLPVGSKVLLAGLFVWLILFLGASKY
YLIRVARNOERALTIVWSGDDKEDLVNNTYVI"

CDS

BASE COUNT 310 a 388 c 464 g 347 t
ORIGIN
Query Match 94.5%; Score 825.4; DB 9; Length 1509;
Best Local Similarity 99.8%; Pred. No. 1e-186;
Matches 837; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 CTCACGCAATATGTTCAATATGAAAGTAATGACCGCCCAAGCACTGAGGCTTG 60
654 CTCACGCAATATGTTCAATATGAAAGTAATGACCGCCCAAGCACTGAGGCTTG 713
61 CCGTCATCTCTCCACGCGTGTGACTTGAAGTGAAGAGAACTCTTCGAAATACCTTCA 120
714 CCGTCATCTCTCCACGCGTGTGACTTGAAGTGAAGAGAACTCTTCGAAATACCTTCA 773
121 CTATGAGAGTGGCCGCGGCAATTAAGAACAGTACCGCTTGAAGAGGCTGATCTCG 180
774 CTATGAGAGTGGCCGCGGCAATTAAGAACAGTACCGCTTGAAGAGGCTGATCTCG 833
181 CTGCTTCGCGAGAGAGAAATCTCCCTGCGCTTGGCTCAAAAGTGTGTCTGGC 240
834 CTGCTTCGCGAGAGAGAAATCTCCCTGCGCTTGGCTCAAAAGTGTGTCTGGC 893
241 GGGGCTGTTCTGATGCTGATCTCTTCTGAGGAGCTCCATGGCTACCTGATCG 300

894 GGGGCTGTTCTGATGCTGATCTCTTCTGAGAGCTCCATGGTACCTGATCGG 953
301 GGTGGACGAGAGAACACAGAGAGTGTGCTGCGACCGTCTGAGAGTCCGGAGATGACAA 360
954 GGTGGACGAGAGAACACAGAGAGTGTGCTGCGACCGTCTGAGAGTCCGGAGATGACAA 1013
361 GAGAGAGCTGTGTGAAGAACACATATGTCCTGAGACCGCTGTGGCCAGAGAGT-69G 419
1014 GAGAGAGCTGTGTGAAGAACACATATGTCCTGAGACCGCTGTGGCCAGAGAGTGGGG 1073
420 AAGGAGGAGAGACATATGTCCTGAGAGCTTTTAAATAGAGGATGAGTCTGATTTAGT 479
1074 AAGGAGGAGAGACATATGTCCTGAGAGCTTTTAAATAGAGGATGAGTCTGATTTAGT 1133
480 GATCATTTAGGAGTGTGAGTCTTCTGTGGAGTAGAGAGGCTGCTCTGCTGTGGA 539
1134 GATCATTTAGGAGTGTGAGTCTTCTGTGGAGTAGAGAGGCTGCTCTGCTGTGGA 1193
540 GGGATGGGTTTCTGTAATCTCTGAGAGGCTCTCTGATGGCTGAGAGTCTGG 599
1194 GGGATGGGTTTCTGTAATCTCTGAGAGGCTCTCTGATGGCTGAGAGTCTGG 1253
600 CAGAGAGGAGAGTGTGCTCTGAGAGTCTCTCTGAGAGTCTCTCTGAGAGTCTCTT 659
1254 CAGAGAGGAGAGTGTGCTCTGAGAGTCTCTCTGAGAGTCTCTCTGAGAGTCTCTT 1313
660 GCTTATGTTGAATTCATGCTCTCTCTGATCAGAGAGTGTGATGCTGCTTCT 719
1314 GCTTATGTTGAATTCATGCTCTCTCTGATCAGAGAGTGTGATGCTGCTTCT 1373
720 TTTGTTGTGTGATTTATGTTTATTAAGTATTAACAAAGCTTTTATATAGATCTG 779
1374 TTTGTTGTGTGATTTATGTTTATTAAGTATTAACAAAGCTTTTATATAGATCTG 1433
780 AAAGAGAGAGAGTAAATGTTCAAGTTAAATTAAGAGGAGGCTCTCTTACATATA 838
1434 AAAGAGAGAGAGTAAATGTTCAAGTTAAATTAAGAGGAGGCTCTCTTACATATA 1492

RESULT 7

AX410783 1544 bp DNA linear PAT 14-JUN-2002
LOCUS AX410783
DEFINITION Sequence 3430 from Patent WO0229103.
ACCESSION AX410783
VERSION AX410783.1 GI:21443488
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 3430 11-Apr-2002;
JOURNAL GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
source 1..1544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. U78095"

BASE COUNT

316 a 394 c 473 g 360 t 1 others
ORIGIN
Query Match 94.5%; Score 825.2; DB 6; Length 1544;
Best Local Similarity 98.5%; Pred. No. 1.2e-186;
Matches 864; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

1 CTCACGCAATATGTTCAATATGAAAGTAATGACCGCCCAAGCACTGAGGCTTG 60
666 CTCACGCAATATGTTCAATATGAAAGTAATGACCGCCCAAGCACTGAGGCTTG 725
61 CCGTCATCTCTCCACGCGTGTGACTTGAAGTGAAGAGAACTCTTCGAAATACCTTCA 120

QY 600 CAGCAGCCCCGAGTTGTTCTCTGATGATGATTTCTTCTCCTCAGGTAGTCTTCTT 659
 DB 1266 CAGCAGCCCCGAGTTGTTCTCTGATGATGATTTCTTCTCCTCAGGTAGTCTTCTT 1325
 QY 660 GCTTATGTTGATTCATGCTGCTC--TTTTCATCAGAGAGATGATGATGCTT 717
 DB 1326 GCTTATGTTGATTCATGCTGCTCCTTCTTCTCAGAGAGATGATGATGCTT 1385
 QY 718 CTTTGTGTTGCTGATTTAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 777
 DB 1386 CTTTGTGTTGCTGATTTAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1445
 QY 778 TGAAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
 DB 1446 TGAAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1505
 QY 838 ATT-CAGCATGCTTCTTCAAAAAAATGATGATGATGATGATGATGATGATGATGAT 873
 DB 1506 ATTCCAGCATGCTTCTTCAAAAAAATGATGATGATGATGATGATGATGATGATGAT 1542

RESULT 9
 AB006534 1512 bp mRNA linear PRI 04-MAR-1998
 LOCUS Homo sapiens mRNA for hepatocyte growth factor activator inhibitor
 DEFINITION type 2, complete cds.
 ACCESSION AB006534
 VERSION AB006534.1 GI:2924619
 KEYWORDS hepatocyte growth factor activator inhibitor type 2.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens CDNA to mRNA.
 Mammalia; Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 AUTHORS Kawaguchi, T., Qin, L., Shimomura, T., Kondo, J., Masumoto, K.,
 Denda, K., and Kitamura, N.
 TITLE Purification and cloning of hepatocyte growth factor activator
 JOURNAL inhibitor type 2, a Kunitz-type serine protease inhibitor
 MEDLINE J Biol Chem. 272 (44), 27558-27564 (1997)
 REFERENCE 98010584
 JOURNAL 2 (bases 1 to 1512)
 TITLE Direct Submission
 AUTHORS Submitted (07-AUG-1997) Kimiotsu Denda, Tokyo Institute of
 JOURNAL Technology, Department of Life Science, 4259 Nagatsuta, Midori-ku,
 Yokohama, Kanagawa 227, Japan (E-mail: kenda@bio.itech.ac.jp,
 Tel:45-924-5702, Fax:45-924-5771)
 FEATURES
 source
 1. .1512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 313..1071
 CDS
 /codon_start=1
 /product="hepatocyte growth factor activator inhibitor
 type 2"
 /protein_id="BA25024.1"
 /db_xref="GI:2924620"
 /translation="MAQLCGLRSPAFLLIGSLISGLAADERSIHDFCLVSKV
 GRCASMPWYNYTDSQCLPYGGCDNSNNYTKRECKCATYTNAGDGLATS
 RNAADSVPASAPRQDSHSDHSENYEYCTANAVTPCHASFPKPYDVERNSCN
 FIVGCRGKNSYRSEACMLRCFROENPPLFSGKVVLAGLFWVILFLGASKV
 YLRVARRRQEPALRTWSSGDDKDLVKNYTVL"
 BASE COUNT 1487 1492
 polyA signal 291 a 396 c 471 g 354 t
 ORIGIN
 Query Match 92.7%; Score 809.4; DB 9; Length 1512;
 Best Local Similarity 99.6%; Pzed. No. 6.9e-183;
 Matches 832; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 QY 1 CTCGAGCATATGTTCAACTATGAGATATGCAAGCCCAAGCAGTCACTGGGCTTG 60
 DB 676 CTCGAGCATATGTTCAACTATGAGATATGCAAGCCCAAGCAGTCACTGGGCTTG 737

QY 61 CCGTGATCTCTTCCACGCTGTAATTGACGTGAGAGAACTCCTGCATTAATTCTTAT 120
 DB 738 CCGTGATCTCTTCCACGCTGTAATTGACGTGAGAGAACTCCTGCATTAATTCTTAT 797
 QY 121 CTATGAGAGCTGCGGGGCAATAGAAACGCTCTGAGAGAGCTCGATGCTCG 180
 DB 798 CTATGAGAGCTGCGGGGCAATAGAAACGCTCTGAGAGAGCTCGATGCTCG 857
 QY 181 CTCTTCCGACAGAGAGAAATCTCCCTGCGCTTGGCTCAAGGTGCTGCTGCG 240
 DB 858 CTCTTCCGACAGAGAGAAATCTCCCTGCGCTTGGCTCAAGGTGCTGCTGCG 917
 QY 241 GGGGCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 918 GGGGCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 977
 QY 301 GGTGACACGAGAGAAACAGAGAGCTGCTGCGCAAGCTGAGAGCTCCGAGATGACAA 360
 DB 978 GGTGACACGAGAGAAACAGAGAGCTGCTGCGCAAGCTGAGAGCTCCGAGATGACAA 1037
 QY 361 GAGACAGCTGCTGAGAGAAACATATGCTGCTGAGACCGCTGCGCAAGAGACT-AGG 419
 DB 1038 GAGACAGCTGCTGAGAGAAACATATGCTGCTGAGACCGCTGCGCAAGAGACTGAGG 1097
 QY 420 AAGGAGAGGAGAGCTATGCTGAGCTTTTTPAATAGAGAGATGACTCGATTTGACT 479
 DB 1098 AAGGAGAGGAGAGCTATGCTGAGCTTTTTPAATAGAGAGATGACTCGATTTGACT 1157
 QY 480 GATCATTAAGGCTGAGAGCTGCTGCTGCTGAGAGAGAGAGAGCTGCTGCTGCTGAG 539
 DB 1158 GATCATTAAGGCTGAGAGCTGCTGCTGCTGAGAGAGAGAGAGCTGCTGCTGCTGAG 1217
 QY 540 GGGATGAGTTGCTTGAATAATCTCTAGAGAGCTCTCTGCTGAGAGAGAGAGAGAG 599
 DB 1218 GGGATGAGTTGCTTGAATAATCTCTAGAGAGCTCTCTGCTGAGAGAGAGAGAGAG 1277
 QY 600 CAGCAGCCCCGAGTTGTTCTCTGATGATGATTTCTTCTCCTCAGGTAGTCTTCTT 659
 DB 1278 CAGCAGCCCCGAGTTGTTCTCTGATGATGATTTCTTCTCCTCAGGTAGTCTTCTT 1337
 QY 660 GCTTATGTTGATTCATGCTGCTCCTTCTCCTCAGAGAGTATGATGATGATGATGAT 719
 DB 1338 GCTTATGTTGATTCATGCTGCTCCTTCTCCTCAGAGAGTATGATGATGATGATGAT 1397
 QY 720 TTTGTTGCTGATTTATGCTTTTATGATTAACAAAG-TTTTATTTACATTTCT 778
 DB 1398 TTTGTTGCTGATTTATGCTTTTATGATTAACAAAGTTTATTTATTTACATTTCT 1457
 QY 779 GAAAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 833
 DB 1458 GAAAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1512

RESULT 10
 AX302543 1564 bp DNA linear PAT 30-NOV-2001
 LOCUS AX302543
 DEFINITION Sequence 61 from Patent WO01/5177.
 ACCESSION AX302543
 VERSION AX302543.1 GI:17383081
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 AUTHORS Morin, P. J., Sherman-Baust, C. A., Pizer, E. S., and Hough, C. D.
 TITLE Tumor makers in ovarian cancer
 JOURNAL Patent: WO 01/5177-A 61 11-OCT-2001;
 THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US)
 FEATURES
 source
 1. .1564
 /organism="Homo sapiens"

```

/d_b_xref="taxon:9606"
BASE COUNT      297 a      419 c      491 g      357 t
ORIGIN
Query Match      92.3%; Score 805.6; DB 6; Length 1564;
Best Local Similarity 99.3%; Pred. No. 5,6e-182;
Matches 830; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 CTCACGCGATATGTTCAACTATGAAATATGCAACCCGCAACCGATCTCTGGGCTTGG 60
DB 730 CTCACGCGATATGTTCAACTATGAAATATGCAACCCGCAACCGATCTCTGGGCTTGG 789
QY 61 CCGTGATCTCTCCGAGCGTGTGACTTGTGACGGTGGAGAGAACTCCGTGCAATATCTCAT 120
DB 790 CCGTGATCTCTCCGAGCGTGTGACTTGTGACGGTGGAGAGAACTCTGCAATATCTCAT 849
QY 121 CTATGAGAGCTCCGCGGCGCAATAGAAACAGCTACCGCTCTGAGAGAGCTGCAATGCTCCG 180
DB 850 CTATGAGAGCTCCGCGGCGCAATAGAAACAGCTACCGCTCTGAGAGAGCTGCAATGCTCCG 909
QY 181 CTGCTTCGCGCAGAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 910 CTGCTTCGCGCAGAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
QY 241 GGGGCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 970 GGGGCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1029
QY 301 GGTGGCAGGAGAGAAACAGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 1030 GGTGGCAGGAGAGAAACAGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089
QY 361 GGAAGCAGCTGTGTAAGAAACACATATGCTGTGACCGCTGTGTGCTGCTGCTGCTGCTGCT 419
DB 1090 GGAAGCAGCTGTGTAAGAAACACATATGCTGTGACCGCTGTGTGCTGCTGCTGCTGCTGCT 1149
QY 420 AAGGAGAGGAGACATATGATGAGCTTTTAAATAGAGAGATGATGATGATGATGATGATG 479
DB 1150 AAGGAGAGGAGACATATGATGAGCTTTTAAATAGAGAGATGATGATGATGATGATGATG 1209
QY 480 GATCATTAAGGGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539
DB 1210 GATCATTAAGGGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1269
QY 540 GGGATGGGATTTGCTTTGAAATCTCTAGAGAGCTCTCTCTGCAATGAGCTGAGTCTGG 599
DB 1270 GGGATGGGATTTGCTTTGAAATCTCTAGAGAGCTCTCTCTGCAATGAGCTGAGTCTGG 1329
QY 600 CAGCAGCCCCGAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
DB 1330 CAGCAGCCCCGAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1389
QY 660 GCTATATGTAATTTCCATGCTCTCTCTGATCAGAGAGATGATGATGATGATGATGATGAT 719
DB 1390 GCTATATGTAATTTCCATGCTCTCTCTGATCAGAGAGATGATGATGATGATGATGATGAT 1449
QY 720 TTTGTTGCTGATTTATGATTTTAAATAGAAAGATTTTAAATAGAAATTTTAAATAGAAAT 779
DB 1450 TTTGTTGCTGATTTATGATTTTAAATAGAAAGATTTTAAATAGAAATTTTAAATAGAAAT 1508
QY 780 AAAGAGAGAAAGTAAATATGATCAAGTTAAATAAAGAGAGAGCTTCCCTCTTAAAT 835
DB 1509 AAAGAGAGAAAGTAAATATGATCAAGTTAAATAAAGAGAGAGCTTCCCTCTTAAAT 1564

RESULT 11
AF027205      AF027205      1564 bp      mRNA      linear      PRI 09-NOV-1997
LOCUS        Homo sapiens Kunitz-type protease inhibitor (kop) mRNA, complete
DEFINITION   cds.
ACCESSION    AF027205
VERSION      AF027205.1  GI:2598967
KEYWORDS

```

```

SOURCE          Homo sapiens.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 1564)
AUTHORS         Mueller-Pillasch,F., Wallrapp,C., Bartels,K., Varga,G., Fritsch,H.,
                Buechler,M., Adler,G. and Gress,T.M.
TITLE           Cloning of a new Kunitz-type protease inhibitor with a putative
                transmembrane domain overexpressed in pancreatic cancer
JOURNAL         Biochim. Biophys. Acta (1997) In press
REFERENCE       2 (bases 1 to 1564)
AUTHORS         Mueller-Pillasch,F., Wallrapp,C., Bartels,K., Adler,G. and
                Gress,T.M.
TITLE           Direct Submission
JOURNAL         Submitted (29-SEP-1997) Medizinische Klinik, Internal Medicine I,
                Robert-Koch-Str 8, Ulm 89081, Germany
FEATURES        Location/Qualifiers
                source
                1..1564
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="19"
                /map="19q13.1"
                /cell_line="Panc 8988"
                /tissue_type="pancreatic cancer"
                1..1564
                /gene="kop"
                /gene="KOP"
                /note="KOP; contains putative transmembrane domain;
                similar to human placental bikunin; overexpressed in
                pancreatic cancer"
                /codon_start=1
                /product="Kunitz-type protease inhibitor"
                /protein_id="AA84031.1"
                /db_xref="GI:2598968"
                /translation="MAHLGLRSPALFALLIGSLISGLVLAADERSIHDECTSKYV
                GRCSAMPKMYVNTDSCOLFVYGGCGNNNYLTKECLKKCATVENATGLATS
                RNADSSVSPAPRRODSEHSDSMFNEYEYCTANAVGPCRASFPWYEDVSRSCNN
                RLYGCGKNNYSRSEACMLRCFQDENPLPLGSRVYVLAGLFWVLILPLGASVY
                YLIRVARNRERALRTVWSSGHDKQLKNTYV"

BASE COUNT      297 a      419 c      491 g      357 t
ORIGIN
Query Match      92.3%; Score 805.6; DB 9; Length 1564;
Best Local Similarity 99.3%; Pred. No. 5,6e-182;
Matches 830; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 CTCACGCGATATGTTCAACTATGAAATATGCAACCCGCAACCGATCTCTGGGCTTGG 60
DB 730 CTCACGCGATATGTTCAACTATGAAATATGCAACCCGCAACCGATCTCTGGGCTTGG 789
QY 61 CCGTGATCTCTCCGAGCGTGTGACTTGTGACGGTGGAGAGAACTCCGTGCAATATCTCAT 120
DB 790 CCGTGATCTCTCCGAGCGTGTGACTTGTGACGGTGGAGAGAACTCCGTGCAATATCTCAT 849
QY 121 CTATGAGAGCTCCGCGGCGCAATAGAAACAGCTACCGCTCTGAGAGAGCTGCAATGCTCCG 180
DB 850 CTATGAGAGCTCCGCGGCGCAATAGAAACAGCTACCGCTCTGAGAGAGCTGCAATGCTCCG 909
QY 181 CTGCTTCGCGCAGAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 910 CTGCTTCGCGCAGAGAGAAATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
QY 241 GGGGCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 970 GGGGCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1029
QY 301 GGTGGCAGGAGAGAAACAGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 1030 GGTGGCAGGAGAGAAACAGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089
QY 361 GGAAGCAGCTGTGTAAGAAACACATATGCTGTGACCGCTGTGTGCTGCTGCTGCTGCTGCT 419

```


LOCUS AC011479 169660 bp DNA linear PRI 14-JUL-2002
 DEFINITION Homo sapiens chromosome 19 clone CTB-10215, complete sequence.
 AC011479
 VERSION AC011479.6 GI:9954635
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 169660)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 169660)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 169660)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 169660)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Aug 31, 2000 this sequence version replaced gi:771511.
 Draft Sequence Produced by DOE Joint Genome Institute
 COMMENT www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.8% of Sequence;
 Estimated Total Number of Errors is 1.2.
 STS Content:
 SHGC-141352 G63443
 Location/Qualifiers
 source 1. 169660
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /clone="CTB-10215"
 BASE COUNT 39214 a 45115 c 44847 g 40484 t
 ORIGIN
 Query Match 71.1%; Score 621; DB 9; Length 169660;
 Best Local Similarity 94.1%; Pred. No. 1.8e-137;
 Matches 656; Conservative 0; Mismatches 40; Indels 1; Gaps 1;
 Oy 160 TGAGGAGGCTCGATGCTCCGCTCTTCCGACAGAGAGATCCCTCCCTGCTG 219
 Db 27605 TGGTGAAGCGCACTCGGCTGCACTCCCTTCCTGAGCCGCTCTCTC 27546
 Oy 220 CTGAAGGCTGCTGCTTCTGCGGCGGCTGCTGATGCTGATCCCTTCTCGGAGC 279
 Db 27545 GTCCCTAGTGTGCTGCTGCGGCGGCTGCTGATGCTGATCCCTTCTCGGAGC 27486
 Oy 280 CTCATGCTCTACTGATCCGGGTGCGACGAGACGAGAGCGTCCGCGACCCCT 339
 Db 27485 CTCATGCTCTACTGATCCGGGTGCGACGAGACGAGAGCGTCCGCGACCCCT 27426
 Oy 340 CTGAGCTCCGAGATGACAGAGACAGCTGTGAAGAACATATGCTCTGTACCGCC 399
 Db 27425 CTGAGCTCCGAGATGACAGAGACAGCTGTGAAGAACATATGCTGTACCGCC 27366
 Oy 400 CTGTCCGCAAGAGACT GGGAGGGAGGGAGACATATGTGTGACCTTTTAAATAGA 458
 Db 27365 CTGTCCGCAAGAGACTGTGGAGAGGGAGACATATGTGTGACCTTTTAAATAGA 27306
 Oy 459 GGGATTGACTCGGATTGAGTATGATATTAGAGGCTGATGCTCTCTCGGAGGTAGA 518
 Db 27305 GGGATTGACTCGGATTGAGTATGATATTAGAGGCTGATGCTCTCTCGGAGGTAGA 27246

Oy 519 CGAGTCTCTCCGCTGCTGCGAGGAGATGGCTTCTTGAGAAATCCTTGAAGCTCTTC 578
 Db 27245 CGAGTCTCTCCGCTGCTGCGAGGAGATGGCTTCTTGAGAAATCCTTGAAGCTCTTC 27186
 Oy 579 CTGCGATGGCTCGAGCTGCGACGAGCCCGAGTGTTCCTCGCTGATGATTTCTT 638
 Db 27185 CTGCGATGGCTCGAGCTGCGACGAGCCCGAGTGTTCCTCGCTGATGATTTCTT 27126
 Oy 639 CCCCAGGAGAGTTCCTTGTGCTATGTTGAATCCATTCCTCTTCTCATCAGAGA 698
 Db 27125 CCCCAGGAGAGTTCCTTGTGCTATGTTGAATCCATTCCTCTTCTCATCAGAGA 27066
 Oy 699 AGTGAATGTAATCGTTCCTTGTGCTATGTTGAATGATTTTATGATTAACAA 758
 Db 27065 AGTGAATGTAATCGTTCCTTGTGCTATGTTGAATGATTTTATGATTAACAA 27006
 Oy 759 AAGTTTATTTAGCATTTCTGAAAGAGAAAGTAAATGTAAGTTTATTAAGAGG 818
 Db 27005 AAGTTTATTTAGCATTTCTGAAAGAGAAAGTAAATGTAAGTTTATTAAGAGG 26946
 Oy 819 GCCTCCCTTTAGAAATTAATTCAGCATGCTTC 855
 Db 26945 GCCTCCCTTTAGAAATTAATTCAGCATGCTTC 26909
 RESULT 14
 AC083794/c 64620 bp DNA linear HTG 30-SEP-2000
 LOCUS AC083794/c
 DEFINITION Homo sapiens chromosome 4 clone RP11-140B3 map 4, LOW-PASS SEQUENCE
 SAMPLING.
 AC083794 64620 bp DNA linear HTG 30-SEP-2000
 VERSION AC083794.1 GI:10440701
 KEYWORDS HTG; HTGS PHASED.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 64620)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Homo sapiens chromosome 4, clone RP11-140B3
 REFERENCE 2 (bases 1 to 64620)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
 Bouckgailler, B., Brown, A., Burkett, G., Campoliano, A., Castle, A.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
 Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glende, S., Goyette, M.,
 Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
 Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRoque, K.,
 Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, C., Liu, G.,
 Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Meldrum, J., Menus, L., Mhova, T., Mlenga, V.,
 Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
 O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Ribback, M., Riley, R.,
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodores, J.,
 Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zaitoun, J.,
 Zimmerman, A. and Zody, M.
 Direct Submission
 Submitted (30-SEP-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/kw/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L1254
Center Clone name: 140_E_3

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 715 814: contig of 714 bp in length
* 815 1522: contig of 708 bp in length
* 1523 1622: gap of 100 bp
* 1623 2347: contig of 725 bp in length
* 2348 2447: gap of 100 bp
* 2448 3182: contig of 735 bp in length
* 3183 3282: gap of 100 bp
* 3283 4020: contig of 738 bp in length
* 4021 4120: gap of 100 bp
* 4121 4848: contig of 728 bp in length
* 4849 4948: gap of 100 bp
* 4949 5677: contig of 729 bp in length
* 5678 5777: gap of 100 bp
* 5778 6523: contig of 746 bp in length
* 6524 6623: gap of 100 bp
* 6624 7362: contig of 739 bp in length
* 7363 7462: gap of 100 bp
* 7463 8130: contig of 728 bp in length
* 8131 8290: gap of 100 bp
* 8291 9000: contig of 710 bp in length
* 9001 9100: gap of 100 bp
* 9101 9825: contig of 725 bp in length
* 9826 9925: gap of 100 bp
* 9926 10646: contig of 721 bp in length
* 10647 10746: gap of 100 bp
* 10747 11461: contig of 715 bp in length
* 11462 11561: gap of 100 bp
* 11562 12292: contig of 731 bp in length
* 12293 12392: gap of 100 bp
* 12393 13135: contig of 743 bp in length
* 13136 13235: gap of 100 bp
* 13236 13959: contig of 724 bp in length
* 13960 14059: gap of 100 bp
* 14060 14795: contig of 736 bp in length
* 14796 14895: gap of 100 bp
* 14896 15642: contig of 747 bp in length
* 15643 15742: gap of 100 bp
* 15743 16462: contig of 720 bp in length
* 16463 16562: gap of 100 bp
* 16563 17306: contig of 744 bp in length
* 17307 17406: gap of 100 bp
* 17407 18117: contig of 711 bp in length
* 18118 18217: gap of 100 bp
* 18218 18945: contig of 728 bp in length
* 18946 19045: gap of 100 bp
* 19046 19761: contig of 736 bp in length
* 19762 19861: gap of 100 bp
* 19862 20605: contig of 724 bp in length
* 20606 20705: gap of 100 bp
* 20706 21446: contig of 741 bp in length
* 21447 21546: gap of 100 bp
* 21547 22282: contig of 736 bp in length
* 22283 22382: gap of 100 bp
* 22383 23165: contig of 743 bp in length
* 23126 23225: gap of 100 bp
* 23226 23947: contig of 722 bp in length

* 23948 24047: gap of 100 bp
* 24048 24776: contig of 729 bp in length
* 24777 24876: gap of 100 bp
* 24877 25594: contig of 718 bp in length
* 25595 25694: gap of 100 bp
* 25695 26441: contig of 747 bp in length
* 26442 26541: gap of 100 bp
* 26542 27283: contig of 742 bp in length
* 27284 27383: gap of 100 bp
* 27384 28121: contig of 738 bp in length
* 28122 28221: gap of 100 bp
* 28222 28957: contig of 736 bp in length
* 28958 29057: gap of 100 bp
* 29058 29808: contig of 751 bp in length
* 29809 29908: gap of 100 bp
* 29909 30651: contig of 743 bp in length
* 30652 30751: gap of 100 bp
* 30752 31508: contig of 757 bp in length
* 31509 31608: gap of 100 bp
* 31609 32337: contig of 729 bp in length
* 32338 32437: gap of 100 bp
* 32438 33175: contig of 738 bp in length
* 33176 33275: gap of 100 bp
* 33276 34002: contig of 726 bp in length
* 34002 34101: gap of 100 bp
* 34102 34799: contig of 698 bp in length
* 34800 34899: gap of 100 bp
* 34900 35623: contig of 724 bp in length
* 35624 35723: gap of 100 bp
* 35724 36453: contig of 730 bp in length
* 36454 36553: gap of 100 bp
* 36554 37271: contig of 718 bp in length
* 37272 37371: gap of 100 bp
* 37372 38098: contig of 727 bp in length
* 38099 38198: gap of 100 bp
* 38199 38906: contig of 708 bp in length
* 38907 39006: gap of 100 bp
* 39007 39719: contig of 713 bp in length
* 39720 39819: gap of 100 bp
* 39820 40520: contig of 701 bp in length
* 40521 40620: gap of 100 bp
* 40621 41363: contig of 743 bp in length
* 41364 41463: gap of 100 bp
* 41464 42184: contig of 721 bp in length
* 42185 42284: gap of 100 bp
* 42285 43030: contig of 746 bp in length
* 43031 43130: gap of 100 bp
* 43131 43865: contig of 735 bp in length
* 43866 43965: gap of 100 bp
* 43966 44706: contig of 741 bp in length
* 44707 44806: gap of 100 bp
* 44807 45521: contig of 715 bp in length
* 45522 45621: gap of 100 bp
* 45622 46341: contig of 720 bp in length
* 46342 46441: gap of 100 bp
* 46442 47166: contig of 725 bp in length
* 47167 47266: gap of 100 bp
* 47267 47981: contig of 715 bp in length
* 47982 48081: gap of 100 bp
* 48082 48817: contig of 736 bp in length
* 48818 48917: gap of 100 bp
* 48918 49635: contig of 718 bp in length
* 49636 49735: gap of 100 bp
* 49736 50462: contig of 727 bp in length
* 50463 50562: gap of 100 bp
* 50563 51287: contig of 725 bp in length
* 51288 51387: gap of 100 bp
* 51388 52107: contig of 720 bp in length
* 52108 52207: gap of 100 bp
* 52208 52943: contig of 736 bp in length
* 52944 53043: gap of 100 bp
* 53044 53774: contig of 731 bp in length
* 53775 53874: gap of 100 bp

* 53875 54611: config of 737 bp in length
 * 54612 54711: gap of 100 bp
 * 54712 55529: config of 718 bp in length
 * 55530 55529: gap of 100 bp
 * 55530 56244: config of 715 bp in length
 * 56245 56344: gap of 100 bp
 * 56345 57090: config of 746 bp in length
 * 57091 57190: gap of 100 bp
 * 57191 57904: config of 714 bp in length

Query Match 68.2%; Score 595.2; DB 2; Length 64620;
 Best Local Similarity 96.1%; Pred. No. 2.2e-131;
 Matches 620; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

QY 202 TCCCTCCCTGCCCCCTTGGCTCAAGGAGTGTCTCTGCGGAGCTTCTGATGAGTGT 261
 Db 3959 TCCGAGGCCCTCTCTCTGCTCAGTGTGTCTCTGCGGAGCTTCTGATGAGTGT 3900
 QY 262 GATCCTCTTCTGAGGAGCTCCATGATCTACCTGATCCGAGGAGCAAGGAGCAAGCA 321
 Db 3889 GATCCTCTTCTGAGGAGCTCCATGATCTACCTGATCCGAGGAGCAAGGAGCAAGCA 3840
 QY 322 GCGTCCCTGCGCACCGTCTGAGCTCCGAGATGACAAGAGCAAGCTGTGAGAAACAC 381
 Db 3839 GCGTCCCTGCGCACCGTCTGAGCTCCGAGATGACAAGAGCAAGCTGTGAGAAACAC 3780
 QY 382 AATATCTGTGACCGCCCTGTCCCAAGAGACT-GGAGAGGAGGAGAGACTATGTGT 440
 Db 3779 AATATCTGTGACCGCCCTGTCCCAAGAGACTGAGGAGGAGAGAGACTATGTGT 3720
 QY 441 GAGCTTTTAAATAGAGGATGACTCGATTTGAGATCACTAGGAGCTGAGGCTGTG 500
 Db 3719 GAGCTTTTAAATAGAGGATGACTCGATTTGAGATCACTAGGAGCTGAGGCTGTG 3660
 QY 501 TTCTCTGAGAGTGAAGAGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 560
 Db 3659 TTCTCTGAGAGTGAAGAGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3600
 QY 561 TCCCTAGAGAGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
 Db 3599 TCCCTAGAGAGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3540
 QY 621 TCGCTAGAGATTTCTTCTCCAGTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
 Db 3539 TCGCTAGAGATTTCTTCTCCAGTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480
 QY 681 CTCTTTTCTGATCAAGAGTGAATGAGATGCTTCTTGTGCTGATTTATGCT 740
 Db 3479 CTCTTTTCTGATCAAGAGTGAATGAGATGCTTCTTGTGCTGATTTATGCT 3420
 QY 741 TTTTAAAGTAAACAAAGTTTTTTTATAGATCTGGAAGAGAGAGTAAATGTA 800
 Db 3419 TTTTAAAGTAAACAAAGTTTTTTTATAGATCTGGAAGAGAGAGTAAATGTA 3360
 QY 801 CAAGTTAAATAAAGGAGGCTTCCCTTAGATTAATTTGAGC 845
 Db 3359 CAAGTTAAATAAAGGAGGCTTCCCTTAGATTAATTTGAGC 3315

RESULT 15
 AR069031 759 bp DNA linear PAT 29-SEP-1999
 LOCUS AR069031
 DEFINITION Sequence 4 from patent US 5854396.
 ACCESSION AR069031
 VERSION AR069031.1 GI:6001238
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.

REFERENCE 1 (bases 1 to 759)
 AUTHORS Shimomura T., Kawaguchi T. and Kitamura N.
 TITLE Protein, DNA coding for same and method of producing the protein
 JOURNAL Patent: US 5854396-A 4 29-DEC-1999;

FEATURES
 source Location/Qualifiers
 BASE COUNT 152 a 213 c 236 g 158 t
 ORIGIN

Query Match 45.1%; Score 394; DB 6; Length 759;
 Best Local Similarity 100.0%; Pred. No. 1.5e-83;
 Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACGCAATATGTTCAACTATGAAAGTACGACCGCAAGCAGTCACTGGGCTTG 60
 Db 366 CTCACGCAATATGTTCAACTATGAAAGTACGACCGCAAGCAGTCACTGGGCTTG 425
 QY 61 CCGTGCATCTCTTCCACGCTGTGACTTGTAGGTGAGAGAACTCTGCATTAATCTTAT 120
 Db 426 CCGTGCATCTCTTCCACGCTGTGACTTGTAGGTGAGAGAACTCTGCATTAATCTTAT 485
 QY 121 CTATGAGAGCTGCGGGGCAATTAAGACGCTACCGCTGAGAGAGGCTGATGCTCG 180
 Db 486 CTATGAGAGCTGCGGGGCAATTAAGACGCTACCGCTGAGAGAGGCTGATGCTCG 545
 QY 181 CTGCTTCCGCGCAGAGAGAACTCTCCCTGCGCTGCTCAAGAGTGTGCTGCTGCTG 240
 Db 546 CTGCTTCCGCGCAGAGAGAACTCTCCCTGCGCTGCTCAAGAGTGTGCTGCTGCTG 605
 QY 241 GAGGCTGTGCTGATGCTGTTGATCTCTTCTGAGAGCTCCATGCTTACTGATCG 300
 Db 606 GAGGCTGTGCTGATGCTGTTGATCTCTTCTGAGAGCTCCATGCTTACTGATCG 665
 QY 301 GGTGGCAGAGAGAGAACTGAGAGAGTCCCTGCGCAGCTGAGAGCTCCGAGATGACA 360
 Db 666 GGTGGCAGAGAGAGAACTGAGAGAGTCCCTGCGCAGCTGAGAGCTCCGAGATGACA 725
 QY 361 GGAGCAGCTGTGAGAGACATATGCTCTGTGA 394
 Db 726 GGAGCAGCTGTGAGAGACATATGCTCTGTGA 759

Search completed: January 10, 2003, 02:42:33
 Job time: 2223.98 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:42:26 ; Search time 241.933 Seconds

(without alignments)
8126.187 Million cell updates/sec

Title: US-09-825-682A-56

Perfect score: 873
Sequence: 1 ctccagcgatattctcaact.....tcaaaaaaaaaaaaaaa 873

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: N Geneseq 10102.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862	98.7	1610	19	AAV33063 Human tissue factor
2	843	96.6	1527	24	AA862308 CDNA sequence #95
3	838	96.0	1696	24	AB055028 Human ovarian anti
4	836.4	95.8	4334	24	ABK34780 Human CDNA for nov
5	829	95.0	1530	21	AA470392 Human placental bi
6	829	95.0	1628	20	AA477551 Human placental tumo
7	825.4	94.5	1578	22	AAH57435 Human intestine ce
8	825.2	94.5	1544	18	AA790731 Human placental bi
9	825.2	94.5	1544	24	ABK84435 Human CDNA differe

10	825.2	94.5	1544	24	ABN96932
11	823.6	94.3	1544	24	AA470406
12	808	92.6	1528	18	AA790733
13	805.6	92.3	1564	22	AB485112
14	758.2	86.8	1649	21	AACT8030
15	425.4	48.7	597	21	AA416487
16	394	45.1	759	18	AA416439
17	355.4	40.7	782	18	AA790734
18	355.4	40.7	782	18	AA790733
19	344.4	39.5	362	22	AA502014
20	316.4	36.2	320	22	AA547235
21	316.4	36.2	320	22	AA547235
22	312.2	35.8	331	24	ABN95550
23	312.2	35.8	331	24	ABN95550
24	310	35.5	310	21	AA470366
25	310	35.5	310	21	AA470366
26	309.4	35.4	708	21	AA470364
27	309.4	35.4	708	21	AA470364
28	288.4	33.0	320	24	AA790732
29	273	31.3	285	23	AA573552
30	262	30.0	286	23	AA573552
31	253.8	29.1	287	23	AA58027
32	242.2	27.7	736	20	AA58647
33	242.2	27.3	736	20	AA58647
34	238.4	27.3	241	22	AA473118
35	226.4	25.9	287	23	AA547318
36	213.8	24.5	234	16	AA72829
37	200	22.9	200	22	AA72829
38	186.2	21.3	396	24	AAK29657
39	186.2	21.3	396	24	AAK29657
40	186.2	21.3	396	24	AAK29657
41	177	20.3	206	21	AA470371
42	176.4	20.2	207	21	AA470371
43	167.4	19.2	276	21	AA470383
44	167.4	19.2	276	21	AA470383
45	167	19.1	207	21	AA470045

ALIGNMENTS

RESULT 1
ID AAV33063 standard; CDNA; 1610 BP.
XX AAV33063;
AC
XX
XX
DT 06-NOV-1998 (first entry)
XX
DE Human tissue factor pathway inhibitor-3 (TFPI-3) cDNA.
XX
XX Human tissue factor pathway inhibitor-3, TFPI-3; blood clot; sepsis;
XX fibrin clot; coronary occlusion; acute myocardial infarction;
XX thrombolytic; peripheral arterial embolism; inflammatory disease;
XX transplant rejection; anticoagulant; blood transfusion;
XX extracorporeal circulation; dialysis; haemophilia; ss.
XX
XX Homo sapiens.
OS
XX
XX
FH Key
FT 361..119 Location/Qualifiers
FT CDS /*tag= a
FT FT /*tag= b
FT sig_peptide 361..441
FT mat_peptide 442..1116
FT /*tag= c
XX
XX NO9833920-A2.
XX
XX 06-AUG-1998.
XX
XX 27-JAN-1998; 98WO-US01468.

Gene #3430 used to
Human placental bi
Human consensus bi
Human placental bi
Kunitz type 2 Ser
Human cancer assoc
Human colon cancer
Hepatocyte growth
Human placental bi
Human placental bi
Bladder cancer-abb
Human breast cancer
Human breast cancer
Gene #2048 used to
Prostate cancer re
Partial human plac
Partial human plac
EST R74593 DNA. H
Human consensus bi
Human ovarian carc
cDNA #32 encoding
cDNA #703 encoding
cDNA #1323 encoding
Human gene express
Human breast cancer
Human breast cancer
cDNA #1258 encodin
Human gene signatu
Colon adenocarcino
Human ovarian carc
Ovarian carcinoma
Partial human plac
Partial human plac
Human ovarian carc
Ovarian carcinoma
Human ovarian carc

XX 31-JAN-1997; 97US-0036703.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Gentz RL, Hsu T, Ni J, Rosen CA;
 XX WPI; 1998-437473/37.
 DR P-PSDB; AAW70286.
 XX
 PT Isolated tissue factor pathway inhibitor-3 - used to develop
 PT products for treating, e.g. pulmonary embolism, thrombosis, sepsis,
 PT inflammatory disease, transplant rejection or haemophilia
 XX
 PS Claim 4; Fig 1A-1B; 57pp; English.
 CC The present sequence represents a human tissue factor pathway
 CC inhibitor-3 (TFPI-3) cDNA which was isolated from osteoblast cDNA
 CC library. The invention also provides the TFPI-3 protein and
 CC screening methods for identifying agonists and antagonists of TFPI-3.
 CC As TFPI-3 inhibits protease activity, it is claimed to be useful for,
 CC e.g. inhibiting intravascular clotting and preventing the formation
 CC of fibrin clots both in vitro and in vivo, for treating coronary
 CC occlusion with acute myocardial infarction and in the prophylaxis
 CC and treatment of peripheral arterial embolism, for the treatment of
 CC sepsis, inflammatory diseases and transplant rejection. TFPI-3 is also
 CC claimed to be useful as an anticoagulant in blood transfusions,
 CC extracorporeal circulation, and dialysis procedures and in blood
 CC samples for laboratory purposes. The TFPI-3 antagonists are claimed
 CC to be useful for promoting coagulation, e.g. in the treatment of
 CC haemophilia.
 XX
 XX Sequence 1610 BP; 331 A; 417 C; 490 G; 372 T; 0 other;
 SQ
 Query Match 98.7%; Score 862; DB 19; Length 1610;
 Best Local Similarity 99.9%; Pred. No 2.4e-207;
 Matches 873; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 CTCGAGCATATGTTCAACTATGAGAAATATGACCGGCAAGCATCTGAGGCTTGG 60
 DB 726 CTCGAGCATATGTTCAACTATGAGAAATATGACCGGCAAGCATCTGAGGCTTGG 785
 QY 61 CCGTGATCTCTCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 DB 786 CCGTGATCTCTCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 845
 QY 121 CTATGAGAGCTGCGGCGCAATAGAAACAGCTACCGCTCTGAGAGAGCTCTGATCTCG 180
 DB 846 CTATGAGAGCTGCGGCGCAATAGAAACAGCTACCGCTCTGAGAGAGCTCTGATCTCG 905
 QY 181 CTGCTTCCGCGACAGAGAAATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 DB 906 CTGCTTCCGCGACAGAGAAATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 965
 QY 241 GGGGCTGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 DB 966 GGGGCTGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025
 QY 301 GGTGACGAGAGAAACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 DB 1026 GGTGACGAGAGAAACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1085
 QY 361 GGAAGAGCTGCTGAGAAACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
 DB 1086 GGAAGAGCTGCTGAGAAACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1145
 QY 420 AAGGAGAGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
 DB 1146 AAGGAGAGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1205
 QY 480 GATCATATGAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
 DB 1206 GATCATATGAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1265

QY 540 GGAATGGGTTTCCTTGGAAATCCTCTAGAGAGCTCTCTCGATGAGGCTGAGTCTG 599
 DB 1266 GGAATGGGTTTCCTTGGAAATCCTCTAGAGAGCTCTCTCGATGAGGCTGAGTCTG 1325
 QY 600 CAGCAGCCCCGAGTGTCTCTGCTGATGATTTCTTCTCTCCAGGTAGATTTTCTT 659
 DB 1326 CAGCAGCCCCGAGTGTCTCTGCTGATGATTTCTTCTCTCCAGGTAGATTTTCTT 1385
 QY 660 GCTATGTTGAATTCATTCCTGCTTCTTCTGATGATGATGATGATGATGATGATGATGAT 719
 DB 1386 GCTATGTTGAATTCATTCCTGCTTCTTCTGATGATGATGATGATGATGATGATGATGAT 1445
 QY 720 TTGTTGTTGCTGATTTATGATTTTATGATTTTATGATTTTATGATTTTATGATTTCTG 779
 DB 1446 TTGTTGTTGCTGATTTATGATTTTATGATTTTATGATTTTATGATTTTATGATTTCTG 1505
 QY 780 AAGAGAGAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
 DB 1506 AAGAGAGAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1565
 QY 840 TTCAGCATGCTTCAAAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 DB 1566 TTCAGCATGCTTCAAAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1599
 RESULT 2
 AAS62308
 ID AAS62308 strand1: cDNA; 1527 BP.
 XX AAS62308;
 AC
 DT 14-FEB-2002 (first entry)
 XX
 DE cDNA sequence #95 encoding novel human secreted protein.
 XX
 KM Human secreted protein; hyperproliferative disorder; autoimmune disorder;
 KM immune deficiency disorder; blood disorder; inflammatory disorder;
 KM infectious disorder; gene therapy; antimicrobial; hepatotropic;
 KM immunosuppressive; antirheumatic; ss.
 XX
 OS Homo sapiens.
 XX
 PN W02001.77291-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 29-MAR-2001; 2001MO-US10485.
 XX
 PR 06-APR-2000; 2000US-195604P.
 XX
 PA (GENY) GENETICS INST INC.
 XX
 PI Wong GG, Clark HF, Fachtel K, Agoestino MJ, Howes SH, Resnick RJ;
 PI Gulkota K, Graham JR;
 XX
 DR WPI; 2002-010900/01.
 XX
 PT New polynucleotides encoding secreted proteins useful for treating e.g.
 PT asthma, HIV and Crohn's disease -
 XX
 PS Claim 1; Page 126-127; 391pp; English.
 CC The present invention relates to the isolation of novel cDNA sequences
 CC which encode human secreted proteins. The cDNA sequences have been
 CC derived from a variety of human tissues. The invention also provides
 CC a method for producing proteins from these polynucleotide sequences.
 CC The proteins are useful for identifying compounds that modulate their
 CC activity and production, and the cell is also useful for identifying
 CC compounds that modulate expression of the polynucleotide sequences
 CC encoding the secreted proteins. The sequences of the invention are
 CC useful for treating diseases such as hyperproliferative disorders
 CC (e.g. cancer), immune deficiency disorders (e.g. severe combined

CC immunodeficiency (SCID), autoimmune disorders (e.g. multiple
 CC sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory
 CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
 CC The polynucleotide sequences of the invention are also useful in gene
 CC therapy. AAs62214-AAs62838 represent the cDNA sequences of the
 CC invention that encode for novel human secreted proteins.
 CC
 CC

Sequence 1527 BP, 296 A, 395 C, 475 G, 361 T, 0 other;

Query Match 96.6%; Score 843; DB 24; Length 1527;

Best Local Similarity 99.9%; Pred. No. 1,4e-202;

Matches 854; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CTCAGGAGATATTTCAATATGAGAGATCTGACCCGCAACGAGTCACTAGGCGCTTG 60
 DB 673 CTCAGGAGATATTTCAATATGAGAGATCTGACCCGCAACGAGTCACTAGGCGCTTG 732
 QY 61 CCGTGCATCTTCCACGCTGTAATTTGAGAGAGAGAACTCTTCATTAATCTTCAAT 120
 DB 733 CCGTGCATCTTCCACGCTGTAATTTGAGAGAGAGAACTCTTCATTAATCTTCAAT 792
 QY 121 CTATGAGAGCTGCGCGGCAATTAAGAACAGCTACCTCTGAGAGAGCTGATGCTCCG 180
 DB 793 CTATGAGAGCTGCGCGGCAATTAAGAACAGCTACCTCTGAGAGAGCTGATGCTCCG 852
 QY 181 CTGCTTCCGCGACAGAGAGAAATCCCTCCCTTGGCTCAAGGTGCTGCTTCCG 240
 DB 853 CTGCTTCCGCGACAGAGAGAAATCCCTCCCTTGGCTCAAGGTGCTGCTTCCG 912
 QY 241 GGGGCTGTTGATGATGTTGATCTCTTCTGAGAGAGCTCATGCTCATGCTGACCG 300
 DB 913 GGGGCTGTTGATGATGTTGATCTCTTCTGAGAGAGCTCATGCTCATGCTGACCG 972
 QY 301 GGTGACACGAGAGAACAGAGAGCTGCTGCGACCGCTCTGAGAGCTGCGAGATGACA 360
 DB 973 GGTGACACGAGAGAACAGAGAGCTGCTGCGACCGCTCTGAGAGCTGCGAGATGACA 1032
 QY 361 GAGAGAGCTGTTGAAACACATATGCTGCTGACCGCTCTGCGCAAGAGAGCT-GGG 419
 DB 1033 GAGAGAGCTGTTGAAACACATATGCTGCTGACCGCTCTGCGCAAGAGAGCT-GGG 1092
 QY 420 AAGGAGAGGAGAGCTATGCTGAGCTTTTAAATAGAGAGATGATCGGATTTAGG 479
 DB 1093 AAGGAGAGGAGAGCTATGCTGAGCTTTTAAATAGAGAGATGATCGGATTTAGG 1152
 QY 480 GATCATTAGGAGCTGAGCTGTTCTCTGAGAGAGAGAGAGCTGCTTCTGATGCGCA 539
 DB 1153 GATCATTAGGAGCTGAGCTGTTCTCTGAGAGAGAGAGAGCTGCTTCTGATGCGCA 1212
 QY 540 GGATAGGTTTGGTTTGGAAATCTCTAGAGAGCTCTCTCTGCAATGCTGCAATGCTG 599
 DB 1213 GGATAGGTTTGGTTTGGAAATCTCTAGAGAGCTCTCTCTGCAATGCTGCAATGCTG 1272
 QY 600 CAGAGAGCCCGAGTTTCTCTGCTGATGAGATTTCTTCTCCAGAGAGAGATTTCTT 659
 DB 1273 CAGAGAGCCCGAGTTTCTCTGCTGATGAGATTTCTTCTCCAGAGAGAGATTTCTT 1332
 QY 660 GCTATGTTGAATTCATTCCTCTTCTCATACAGAGAGATGATGAGATGCTTCT 719
 DB 1333 GCTATGTTGAATTCATTCCTCTTCTCATACAGAGAGATGATGAGATGCTTCT 1392
 QY 720 TTTGTTGTTCTGATTTATAGTTTCTTAAAGATTAACAAAGTTTATTAAGATCTG 779
 DB 1393 TTTGTTGTTCTGATTTATAGTTTCTTAAAGATTAACAAAGTTTATTAAGATCTG 1452
 QY 780 AAAGAGAGAAAGTAAATGTAACATTAATTAAGAGAGAGCTTCCCTTTAGAAATTAAT 839
 DB 1453 AAAGAGAGAAAGTAAATGTAACATTAATTAAGAGAGAGCTTCCCTTTAGAAATTAAT 1512
 QY 840 TTGAGCATGTGCTTT 854
 DB 1513 TTGAGCATGTGCTTT 1527

RESULT 3

AB055028

ID AB055028 standard; cDNA; 1696 BP.

AC AB055028;

XX 22-AUG-2002 (first entry)

XX Human ovarian antigen HDABR73 cDNA, SEQ ID NO:908.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

XX ovarian cancer; breast cancer; tumour; reproductive system disorder;

XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;

XX PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;

XX inflammatory condition; immune disorder; blood disorder;

XX cardiovascular disorder; respiratory disorder; neurological disorder;

XX gastrotesticular disorder; urinary system disorder; drug screening;

XX gene therapy; chromosome mapping; forensic analysis;

XX antibody preparation; cytostatic; immunomodulatory; neuroprotective;

XX antiinflammatory; gynaecological; reproductive; chromosome 19q13.1;

XX gene; ss.

XX Homo sapiens.

XX MO200200677-A1.

XX 03-JAN-2002.

XX 07-JUN-2001; 2001MO-US18569.

XX 07-JUN-2000; 2000US-209467P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI: 2002-147878/19.

XX DR P-PSDB; ABP41951.

XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,

XX useful in the prevention, treatment and diagnosis of cancer (e.g.

XX ovarian cancer), immune disorders, cardiovascular disorders and

XX neurological diseases -

XX Claim 1; SEQ ID NO 908; 2922bp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-

XX ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also

XX encompasses polypeptides 90% identical and polynucleotides 95% identical

XX to the sequences of the invention. The invention additionally relates to

XX recombinant vectors and host cells comprising human ovarian antigen

XX polynucleotides, antibodies against human ovarian antigens, and the use

XX of ovarian antigen polynucleotides and polypeptides in diagnosing,

XX treating, prognosing or preventing various ovary and/or breast-related

XX disorders. Such conditions include ovarian cancer and breast cancer, and

XX metastatic tumours of ovarian or breast origin, reproductive system

XX disorders (e.g., infertility, disorders of pregnancy, anovulation,

XX polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine

XX disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic

XX shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

XX vaginitis), immune disorders (e.g., congenital and acquired

XX immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),

XX blood-related disorders (e.g., anaemia), cardiovascular disorders,

XX respiratory disorders, neurological disorders, gastrotesticular disorders

XX and urinary system disorders. Ovarian antigen polypeptides and

XX polynucleotides may also be used in screening for compounds which

XX modulate ovarian antigen expression or activity. The polynucleotides may

XX further be used for gene therapy, chromosome mapping, in the

XX identification of individuals and in forensic analysis, and the

XX polypeptides may be used as food additives or to prepare antibodies

XX useful in disease diagnosis, drug targeting and phenotyping. The present

XX sequence represents cDNA encoding a human ovarian antigen of the

CC invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

5Q Sequence 1696 BP; 359 A; 450 C; 510 G; 375 T; 2 other;

Query Match 96.0%; Score 838; DB 24; Length 1696;
Best Local Similarity 98.2%; Pred. No. 2.7e-201;
Matches 858; Conservative 0; Mismatches 15; Indels 1; Gaps 1.

[illegible]

ABK34780/c
ID ABK34780 standard; cDNA; 4334 BP.

AC ABK34780;

DT 08-MAY-2002 (first entry)

Human cDNA for novel secreted protein, SEQ ID 549

KM human / ss; gene; secreted protein; immune deficiency; viral infection;
KM bacterial infection; fungal infection; autoimmune disorder; burn;
KM rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KM diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KM Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KM consultation disorder; haemophilia; inflammatory disorder; ulcer;
KM tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KM lymphoid cell deficiency.

OS Homo sapiens.

PN WO200177290-A2

PD 18-OCT-2001
XX

PF 29-MAR-2001; 2001WO-US10295.
VY

PR 06-APR-2000; 2000US-194941P.
YY

PA (GEMY) GENETICS INST INC
XX

PI Wong GG, Clark HF, Hechtel K, Agostino MJ, Howes SH, Resnick RJ,
PI Gulyukta K, Graham JP;

XX
DB
WBT: 2002-179322/23

XX
DT
Six hundred and twelve.

PT human tissue sources which encode secreted proteins, useful for treating immune deficiencies and disorders such as autoimmune disorders

— 15 —

The invention relates to 625 polynucleotides which have been derived from a variety of human tissue sources and which encode novel secreted proteins, their complements and sequences that hybridize to them. Also included are a vector comprising the polynucleotide, a host cell transformed with the vector, the proteins encoded by the polynucleotides, antibodies that bind to the proteins and identification of modulators of the proteins or the expression of the polynucleotide. The polynucleotides can be used as probes for the identification and isolation of full length cDNA and genomic DNA. The polynucleotides and proteins can also be used as nutritional supplements. The protein is useful in the treatment of various immune deficiencies and disorders such as viral infections, bacterial infections, fungal infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions and conditions (e.g. asthma). They are also useful for treating neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease), liver fibrosis, coagulation disorders (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and tumours. They are also useful for tissue regeneration, for wound healing and in the treatment of burns, incisions and ulcers. The proteins are also useful for regulating hematopoiesis, for treating myeloid or lymphoid cell deficiencies. The present sequence is one of the 625 cDNA sequences encoding a secreted protein.

SQ Sequence 4334 BP; 1204 A; 1018 C; 891 G; 1221 T; 0 other;

Query Match	95.8%	Pred. 836.4	DE 24	Length 4333
Best Local Similarity	98.1%	Pred. No. 1.e-200		
Matches 857	Conservative 16	Indels 1	Gaps 1	
1 CTCACGCGATATGTTCACTATGAGAAATACGCAACCGCCCAAGCAGTCATCGGCGCTTG	60			

29 1 CTCGAGGATATGTTCACTATGAGAATACTGCACCGCCCAACGCACTCACTGGCCTTG 60

```

Db 3542 CTCACGCCATATGTTCACTATGAAATATGACACGCCACGACGATCACTGGGCTTG 3483
QY 61 CCGTGATCTCTCCACGCTGTACTTTGACGTGAGAGGAACCTCTGCAATTAATTAT 120
Db 3482 CCGTGATCTCTCCACGCTGTACTTTGACGTGAGAGGAACCTCTGCAATTAATTAT 3423
QY 121 CTATGAGAGCTGCGGGGCAATTAAGAACAGCTACCGCTCTGAGAGGCTCTGATGCTCG 180
Db 3422 CTATGAGAGCTGCGGGGCAATTAAGAACAGCTACCGCTCTGAGAGGCTCTGATGCTCG 3363
QY 181 CTGCTCCGCGACAGAGGAATCTCCCTGCCCCCTTGAGCTCAAGGTGGTGTCTGAC 240
Db 3362 CTGCTCCGCGACAGAGGAATCTCCCTGCCCCCTTGAGCTCAAGGTGGTGTCTGAC 3303
QY 241 GGGGCTGTCTGTATGATGATCTCTTCTGGAGCTCTCAAGGTGTCTGATCTGATCG 300
Db 3302 GGGGCTGTCTGTATGATGATCTCTTCTGGAGCTCTCAAGGTGTCTGATCTGATCG 3243
QY 301 GGTGACAGAGGAACCAAGAGCTGCTCGGCAACCGCTGAGAGCTCGGAGATGACAA 360
Db 3242 GGTGACAGAGGAACCAAGAGCTGCTCGGCAACCGCTGAGAGCTCGGAGATGACAA 3183
QY 361 GAGCAGCTGTGTAAACACATATGCTGTGACCGCCCTGTGACCAAGAGACT-696 419
Db 3182 GAGCAGCTGTGTAAACACATATGCTGTGACCGCCCTGTGACCAAGAGACTG66 3123
QY 420 AAGGAGGAGAGACTATGTGAGCTTTTAAATAGAGGATTTGACTCGATTGAGT 479
Db 3122 AAGGAGGAGAGACTATGTGAGCTTTTAAATAGAGGATTTGACTCGATTGAGT 3063
QY 480 GATCATTTAGGCTGAGGCTCTTCTCTGGAGCTGAGACGAGCTGCTCTGCTGCTGCA 539
Db 3062 GATCATTTAGGCTGAGGCTCTTCTCTGGAGCTGAGACGAGCTGCTCTGCTGCTGCA 3003
QY 540 GGGATGGGTTCTTGAAGATCTCTGAGAGGCTCTCTCGCATGAGCTGAGCTGAG 599
Db 3002 GGGATGGGTTCTTGAAGATCTCTGAGAGGCTCTCTCGCATGAGCTGAGCTGAG 2943
QY 600 CAGACGCCGAGGTTCTCTCGCATGAGATTTCTTCTCGAGTAGAGTTTCTT 659
Db 2942 CAGACGCCGAGGTTCTCTCGCATGAGATTTCTTCTCGAGTAGAGTTTCTT 2883
QY 660 GCTTATGTTGAATTCATTCGCTCTTCTCTCATCACAAGATGATGTTGATGCTTCT 719
Db 2882 GCTTATGTTGAATTCATTCGCTCTTCTCTCATCACAAGATGATGTTGATGCTTCT 2823
QY 720 TTTGTTGCTGATTTATGTTTAAATTAAGTAACAAGTTTAAATTAAGTCTG 779
Db 2822 TTTGTTGCTGATTTATGTTTAAATTAAGTAACAAGTTTAAATTAAGTCTG 2763
QY 780 AAGAGAGGAAGTAAATGTACAAGTTAAATAAAGGGCTCTCCCTTGAATTAAT 839
Db 2762 AAGAGAGGAAGTAAATGTACAAGTTAAATAAAGGGCTCTCCCTTGAATTAATA 2703
QY 840 TTCAGCATGTCTTCAAAAAA 873
Db 2702 AAAAAA 2669

```

RESULT 5
AAA70392
ID AAA70392 standard; DNA; 1530 BP.

AAA70392;

02-FEB-2001 (first entry)

Human placental bikunin protein coding sequence # 2.

Human; mucociliary dysfunction; mucus; sputum;
Chronic obstructive lung disease; chronic bronchitis; CF; Bronchiectasis;
KW BE; asthma; cystic fibrosis; CF; bacterial infection; placental bikunin;

```

KW Kunitz-type serine protease inhibitor; chronic sinusitis; glue ear; ss.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 316..1059
FT /*tag= a
FT /product= "Human Placental bikunin"
XX
XX W0200037099-A2.
XX
XX 29-JUN-2000.
XX
XX 22-DEC-1999; 99W0-GB04381.
XX
XX 22-DEC-1998; 98US-0218913.
XX
XX 17-NOV-1999; 99US-0441966.
XX
XX (FARB ) BAYER AG.
XX
XX Hall R, Poll CT, Newton BB, Taylor WJA;
XX WPI; 2000-452127/39.
XX
XX P-PSDB; AAB14183.
XX
XX Stimulating mucociliary clearance rate of mucus and sputum in lung
XX airways for treating lung diseases such as cystic fibrosis and
XX bronchitis involves administering a Kunitz-type serine protease
XX inhibitor -
XX
XX Disclosure: Figure 4C; 173pp; English.
XX
XX Mucociliary dysfunction is the inability of ciliated epithelium to clear
XX mucus and sputum in lung airways. Mucociliary dysfunction is a serious
XX complication of chronic obstructive lung diseases such as Chronic
XX Bronchitis (CB), Bronchiectasis (BE), asthma and Cystic Fibrosis (CF).
XX In addition, patients suffering from mucociliary dysfunction are
XX susceptible to secondary bacterial infections. The present sequence is
XX a coding sequence for human placental bikunin protein. The protein
XX encoded by the present sequence is a Kunitz-type serine protease
XX inhibitor protein, which can stimulate the rate of mucociliary clearance
XX of mucus and sputum in lung airways. Therefore, the protein encoded by
XX the present sequence may be used for treating lung diseases such as CF,
XX CB, BE, and chronic sinusitis and glue ear which are caused by retention
XX and accumulation of mucus.
XX
XX Sequence 1530 BP; 320 A; 386 C; 467 G; 354 T; 3 other;
XX
XX Query Match 95.0%; Score 829; DB 21; Length 1530;
XX Best Local Similarity 98.2%; Pred. No. 4, 9e-199;
XX Matches 849; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

```

```

QY 1 CTCAGGATATGTTCACTATGAAATATGACACGCCACGACGATCACTGGGCTTG 60
Db 666 CTCAGGATATGTTCACTATGAAATATGACACGCCACGACGATCACTGGGCTTG 725
QY 61 CCGTGATCTCTCCACGCTGTACTTTGACGTGAGAGGAACCTCTGCAATTAATTAT 120
Db 726 CCGTGATCTCTCCACGCTGTACTTTGACGTGAGAGGAACCTCTGCAATTAATTAT 785
QY 121 CTATGAGAGCTGCGGGGCAATTAAGAACAGCTACCGCTCTGAGAGGCTCTGATGCTCG 180
Db 786 CTATGAGAGCTGCGGGGCAATTAAGAACAGCTACCGCTCTGAGAGGCTCTGATGCTCG 845
QY 181 CTGCTCCGCGACAGAGGAATCTCCCTGCCCCCTTGAGCTCAAGGTGGTGTCTGAC 240
Db 846 CTGCTCCGCGACAGAGGAATCTCCCTGCCCCCTTGAGCTCAAGGTGGTGTCTGAC 905
QY 241 GGGGCTGTCTGTATGATGATCTCTTCTGGAGCTCTCAAGGTGTCTGATCTGATCG 300
Db 906 GGGGCTGTCTGTATGATGATCTCTTCTGGAGCTCTCAAGGTGTCTGATCTGATCG 965
QY 301 GGTGACAGAGGAACCAAGAGCTGCTCGGCAACCGCTGAGAGCTCGGAGATGACAA 360

```


[illegible]

XX 03-OCT-2000; 2000US-237189P.
 PR (GENE-) GENE LOGIC INC.
 PA
 XX
 PI Beazer-Barclay Y, Weissman SK, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity
 XX
 PS Claim 1; SEQ ID No 1006; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome, renal
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection, and
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1544 BP; 316 A; 394 C; 473 G; 360 T; 1 other;

Query Match 94.5%; Score 825.2; DB 24; Length 1544;
 Best Local Similarity 98.5%; Pred. No. 4.5e-198;
 Matches 864; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

QY 1 CTCAGGATATGCTCACTAATGAAGATCTGACCGCAACGACGACTGGGCTTG 60
 DB 666 CTCAGGATATGCTCACTAATGAAGATCTGACCGCAACGACGACTGGGCTTG 725
 QY 61 CCGTGATCTTCCACGCTGTAATCTGACGTGAGAGAACTCTGCATAATTCAT 120
 DB 726 CCGTGATCTTCCACGCTGTAATCTGACGTGAGAGAACTCTGCATAATTCAT 785
 QY 121 CTATGAGGCTGCGGGGCAATAAGAACACCTACCCCTCTGAGAGGCTGATGCTCG 180
 DB 786 CTATGAGGCTGCGGGGCAATAAGAACACCTACCCCTCTGAGAGGCTGATGCTCG 845
 QY 181 CTGCTCCGCGACGAGAGATCTCCCTGCCCCCTTGCTCAAGGTGGTCTGCGC 240

DB 846 CTGCTCCGCGACGAGAAATCTCCCTGCCCTTGAGCTCAAAAGTGCTGTCG 905
 QY 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 DB 906 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 965
 QY 301 GGTGACGACGAGAAACGAGAGCGTCCCTGCGCACCGTCTGAGAGCTCCGAGATGACAA 360
 DB 966 GTGAGACGAGAAACGAGAGCGTCCCTGCGCACCGTCTGAGAGCTCCGAGATGACAA 1025
 QY 361 GAGACAGCTGTGAAAGAACATATGCTCTGAGACCGCTCTGCGCAAGAGACT- 419
 DB 1026 GAGACAGCTGTGAAAGAACATATGCTCTGAGACCGCTCTGCGCAAGAGACTG 1085
 QY 420 AAGGAGGAGAGAACTATGAGAGCTTTTAAATAGAGAGATGAGCTCGATTGAGT 479
 DB 1086 AAGGAGGAGAGAACTATGAGAGCTTTTAAATAGAGAGATGAGCTCGATTGAGT 1145
 QY 480 GATCATTAGAGCTGAGAGCTGTTCTCTGAGAGTAAAGACGCTGCTGCTGAGAG 539
 DB 1146 GATCATTAGAGCTGAGAGCTGTTCTCTGAGAGTAAAGACGCTGCTGAGAG 1205
 QY 540 GGGATGGGTTTCTTTGGAATCTCTAGAGAGCTCTCTGAGAGCTGAGAGCTG 599
 DB 1206 GGGATGGGTTTCTTTGGAATCTCTAGAGAGCTCTCTGAGAGCTGAGAGCTG 1265
 QY 600 CAGCAGCCCGAGTGTCTCTGAGAGAGTCTTTCCCGAGAGTATGATTTCTTT 659
 DB 1266 CAGCAGCCCGAGTGTCTCTGAGAGAGTCTTTCCCGAGAGTATGATTTCTTT 1325
 QY 660 GCTTATGTTGAATTCATTTGCTCT- -TTTTCATCAGAGAGTGTGAGTGTG 717
 DB 1326 GCTTATGTTGAATTCATTTGCTCTCTTTTCTCATCAGAGAGTGTGAGTGTG 1385
 QY 718 CTTTGTGTTGCTGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 777
 DB 1286 CTTTGTGTTGCTGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1445
 QY 778 TGAAAGAGAGAAATGAAATGATGAAATGAAATGAAATGAAATGAAATGAAATG 837
 DB 1446 TGAAAGAGAGAAATGAAATGATGAAATGAAATGAAATGAAATGAAATGAAATG 1505
 QY 838 ATTTT-CAGCATGCTTCAAAAAAATGAAAAA 873
 DB 1506 ATTTTCAAGATGCTTCAAAAAAATGAAAAA 1542

RESULT 10
 ABR96932
 ID ABR96932 standard; DNA, 1544 BP.
 AC ABR96932;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Gene #3430 used to diagnose liver cancer.
 XX
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumour; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 OS Homo sapiens.
 XX
 PN W0200229103-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001WO-US30589.
 XX
 PR 02-OCT-2000; 2000US-237054P.
 XX
 PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX WPI; 2002-426119/45.
 XX
 PT Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample
 XX
 PS Claim 1; SEQ ID NO 3430; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytosolic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1544 BP; 316 A; 394 C; 473 G; 360 T; 1 other;

Query Match 94.5%; Score 825.2; DB 24; Length 1544;

Best Local Similarity 98.5%; Pred. No. 4.5e-198;
 Matches 864; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

QY 1 CTCGACGATGTTTCAATAGTAAGTAATGACACCGCCAAAGCAAGTACGAGCCCTTG 60
 DB 666 CTCGACGATGTTTCAATAGTAAGTAATGACACCGCCAAAGCAAGTACGAGCCCTTG 725
 QY 61 CCGTCATCTCTCCACGCGTGTACTTTGACGTGAGAGAACTCTGCAATTAATCTCAT 120
 DB 726 CCGTCATCTCTCCACGCGTGTACTTTGACGTGAGAGAACTCTGCAATTAATCTCAT 785
 QY 121 CTAATGAGAGCTGCGGGGCAATTAAGAACGCTACCGTCTGAGAGAGCCCTGATGCTCCG 180
 DB 786 CTAATGAGAGCTGCGGGGCAATTAAGAACGCTACCGTCTGAGAGAGCCCTGATGCTCCG 845
 QY 181 CTGCTTCGCCAGAGAGAAATCTCCCTCCCTTGGCTCAAGGTGGTGGTCTGGC 240
 DB 846 CTGCTTCGCCAGAGAGAAATCTCCCTCCCTTGGCTCAAGGTGGTGGTCTGGC 905
 QY 241 GGGGCTGTGTGATGATGATGATCTCTCTGAGAGCCCTCAAGGTCTCACTGATCCG 300
 DB 906 GGGGCTGTGTGATGATGATGATCTCTCTGAGAGCCCTCAAGGTCTCACTGATCCG 965
 QY 301 GGTGACGAGAGAAACGAGAGCGTGCCTGCGACCGTCTGAGCTCCGAGATGACAA 360
 DB 966 GGTGACGAGAGAAACGAGAGCGTGCCTGCGACCGTCTGAGCTCCGAGATGACAA 1025
 QY 361 GAGAGAGCTGTGTAAGAACATATGCTGTGACCCCTCTGCGCAAGAGACT-GGG 419
 DB 1026 GAGAGAGCTGTGTAAGAACATATGCTGTGACCCCTCTGCGCAAGAGACTGGGG 1085
 QY 420 AAGGAGAGGAGACTATGATGAGCTTTTAAATAGAGAGATGACCTGCAATTTAGT 479
 DB 1086 AAGGAGAGGAGACTATGATGAGCTTTTAAATAGAGAGATGACCTGCAATTTAGT 1145
 QY 480 GATCATTAAGGCTGAGGTCTGTTCTCTGAGAGGTAGAGCGGTCTCTGTTGGCA 539
 DB 1146 GATCATTAAGGCTGAGGTCTGTTCTCTGAGAGGTAGAGCGGTCTCTGTTGGCA 1205
 QY 540 GGAATGGGTGCTTGGAAATCCCTAGAGAGCTCTCTCTGCAAGGCGTGGAGTGG 599
 DB 1206 GGAATGGGTGCTTGGAAATCCCTAGAGAGCTCTCTCTGCAAGGCGTGGAGTGG 1265

QY 600 CAGACCCCCGAGTGTCTTCCTGCTGATGATTTCTTCCAGGTAGATTTCTTT 659
 DB 1266 CAGACCCCCGAGTGTCTTCCTGCTGATGATTTCTTCCAGGTAGATTTCTTT 1325
 QY 660 GCTTATGTTGAATTCATGCTTC--TTTTCATCAGAGATGATGTTGAATGCTTT 717
 DB 1326 GCTTATGTTGAATTCATGCTTCCTTCATCAGAGATGATGTTGAATGCTTT 1385
 QY 718 CTTTGTGTTGCTGATTTATGTTTATTAAGTAAACAAAGTTTATTAAGCATTC 777
 DB 1386 CTTTGTGTTGCTGATTTATGTTTATTAAGTAAACAAAGTTTATTAAGCATTC 1445
 QY 778 TGAAGAAGAGAAATGTAATGTAATTAATAAAGGCGCTTCCCTTAGAATA 837
 DB 1446 TGAAGAAGAGAAATGTAATGTAATTAATAAAGGCGCTTCCCTTAGAATA 1505
 QY 838 ATTT-CAGCATGCTTCTCAAAAAAAAAAAAAAAAA 873
 DB 1506 ATTTCCAGCATGCTTCTCAAAAAAAAAAAAAAAAA 1542

RESULT 11

AAA70406
 ID AAA70406 standard; DNA, 1544 BP.

AAA70406;

02-FEB-2001 (first entry)

Human placental bikunin protein coding sequence # 3.

Human mucociliary dysfunction; mucus; sputum;
 chronic obstructive lung disease; chronic bronchitis; CB; Bronchiectasis;
 KB; asthma; cystic fibrosis; CF; bacterial infection; placental bikunin;
 Kunitz-type serine protease inhibitor; chronic sinusitis; glue ear; ss.

Homo sapiens.

Key Location/Qualifiers

CD8 301..1059
 /tag= a
 /product= "human placental bikunin"

W0200037099-A2.

29-JUN-2000.

22-DEC-1999; 99WO-GB04381.

22-DEC-1998; 98US-0218913.

17-NOV-1999; 99US-0441966.

(FARB) BAYER AG.

Hall R, Poll CT, Newton BB, Taylor WCA;

WPI; 2000-452127/39.

P-PSDB; AAB14187.

Stimulating mucociliary clearance rate of mucus and sputum in lung
 always for treating lung diseases such as cystic fibrosis and
 bronchitis involves administering a Kunitz-type serine protease
 inhibitor

Disclosure; Figure 4F; 173bp; English.

Mucociliary dysfunction is the inability of ciliated epithelium to clear
 mucus and sputum in lung airways. Mucociliary dysfunction is a serious
 complication of chronic obstructive lung diseases such as Chronic
 Bronchitis (CB), Bronchiectasis (BE), asthma and Cystic Fibrosis (CF).
 In addition, patients suffering from mucociliary dysfunction are
 susceptible to secondary bacterial infections. The present sequence is
 a coding sequence for human placental bikunin protein. The protein

CC encoded by the present sequence is a Kunitz-type serine protease
 CC inhibitor protein, which can stimulate the rate of mucociliary clearance
 CC of mucus and sputum in lung airways. Therefore, the protein encoded by
 CC the present sequence may be used for treating lung diseases such as CF,
 CC CB, BE, and chronic sinusitis and glue ear which are caused by retention
 CC and accumulation of mucus.

XX
 XX Sequence 1544 BP; 316 A; 393 C; 474 G; 360 T; 1 other;

Query Match 94.3%; Score 823.6; DB 21; Length 1544;
 Best Local Similarity 98.4%; Pred. No. 1.1e-197;
 Matches 863; Conservative 0; Mismatches 10; Indels 4; Gaps 3;

QY 1 CTCACGATATGTTCACTATGAAATATGCAACCCCAACGACGATCGAGCTTG 60
 DB 666 CTCACGATATGTTCACTATGAAATATGCAACCCCAACGACGATCGAGCTTG 725
 QY 61 CCGTGATCTCCCGACCGCTGACTTGTGCGGAGAGAACTCCGCAATATCTCAT 120
 DB 726 CCGTGATCTCCCGACCGCTGACTTGTGCGGAGAGAACTCCGCAATATCTCAT 785
 QY 121 CTATGAGAGCTGCGGAGCAATAGAAACAGCTACCGCTGTGAGAGAGCTGATGCTCG 180
 DB 786 CTATGAGAGCTGCGGAGCAATAGAAACAGCTACCGCTGTGAGAGAGCTGATGCTCG 845
 QY 181 CTGCTCCGCAAGAGAAATCTCCCTCCCTGCTGCTCAAGGTGTGTGCTG 240
 DB 846 CTGCTCCGCAAGAGAAATCTCCCTCCCTGCTGCTGCTCAAGGTGTGTGCTG 905
 QY 241 GGGGCTGTTCCGATGAGTGTGATCTCTCTGAGAGCTCCATGCTTACTGATCCG 300
 DB 906 GGGGCTGTTCCGATGAGTGTGATCTCTCTGAGAGCTCCATGCTTACTGATCCG 965
 QY 301 GGTGACAGAGAGAAACAGAGAGCTGCGCCGCAACCGCTGAGAGCTCCGAGATGACA 360
 DB 966 GGTGACAGAGAGAAACAGAGAGCTGCGCCGCAACCGCTGAGAGCTCCGAGATGACA 1025
 QY 361 GAGACAGCTGTGAAACACATATGTCTGTGACCGCCCTGTGCCAAGAGACT-393 419
 DB 1026 GAGACAGCTGTGAAACACATATGTCTGTGACCGCCCTGTGCCAAGAGACT-393 1085
 QY 420 AAGGAGAGAGAGACTATGTGTGAGCTTTTAAATAGAGGATTTGCTCGATTGAGT 479
 DB 1086 AAGGAGAGAGAGACTATGTGTGAGCTTTTAAATAGAGGATTTGCTCGATTGAGT 1145
 QY 480 GATCATTAGGGCTGAGGCTGTTCTCTGAGAGAGAGAGAGCTGCTGCTGCTGCA 539
 DB 1146 GATCATTAGGGCTGAGGCTGTTCTCTGAGAGAGAGAGAGCTGCTGCTGCTGCA 1205
 QY 540 GGGATGGCTTGTCTTGAAATCCTCTAGAGAGCTCCTCGCATGGCTGAGCTG 599
 DB 1206 GGGATGGCTTGTCTTGAAATCCTCTAGAGAGCTCCTCGCATGGCTGAGCTG 1265
 QY 600 CAGCAGCCCGAGTGTGTTCTCTGCTGATGCAATTTCTTCTCCAGGTAGAGTTTCTTT 659
 DB 1266 CAGCAGCCCGAGTGTGTTCTCTGCTGATGCAATTTCTTCTCCAGGTAGAGTTTCTTT 1325
 QY 660 GCTTATGTTGAATTCATGCTC--TTTTCATCAGAGAGAGAGTGTGAAATGTTT 717
 DB 1326 GCTTATGTTGAATTCATGCTCCTTTCTCTCATCAGAGAGAGTGTGAAATGTTT 1385
 QY 718 CTTTGTGTTGCTGATTTATGTTTAAATAGTAACAATAATTTTATATAGACTTC 777
 DB 1386 CTTTGTGTTGCTGATTTATGTTTAAATAGTAACAATAATTTTATATAGACTTC 1445
 QY 778 TGAAGAGAGAAAGTAAATGTACAAGTTAAATAAAGGGGCTTCCCTTTAGAAATA 837
 DB 1446 TGAAGAGAGAAAGTAAATGTACAAGTTAAATAAAGGGGCTTCCCTTTAGAAATA 1505
 QY 838 ATTT-CAGCATGTCTTCAAAAAAATAAAAA 873
 DB 1506 ATTTCAAGCATGTCTTCAAAAAAATAAAAA 1542

RESULT 12

AAT90733
 ID AAT90733 standard; CDNA; 1528 BP.

XX
 AC AAT90733;

XX
 DT 20-APR-1998 (first entry)

XX
 DE Human consensus bikunin cDNA.

XX
 KM Human; consensus bikunin; inhibition; trypsin; kallikrein;

XX
 KM plasmin; factor Xlla; treatment; prevention; oedema;

XX
 KM inflammation; infection; granulomatosis; multiple sclerosis;

XX
 KM ischaemia; perioperative blood loss; sepsis; shock; fibrosis;

XX
 KM blood coagulation disease; polytrauma; stroke; haemorrhage;

XX
 KM gastric cancer; cervical cancer; metastasis; blood loss; ss.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers

XX
 FT CDS 313..1059

XX
 FT /tag= a

XX
 FT /product= bikunin

XX
 FT unsure 313..315

XX
 FT /tag= b

XX
 FT /note= "encodes Met"

XX
 PA (PARB) BAYER CORP.

XX
 PI Davis G, Delaria KA, Marlor CW, Muller DK, Tamburini PP,

XX
 DR WPI; 1997-470876/43.

XX
 DR P-PSDB; AAW30044.

XX
 PT New human placental bikunin - used to inhibit kallikrein, trypsin

XX
 PT etc. in treatment of oedema, multiple sclerosis, fibrosis, or

XX
 PT perioperative blood loss

XX
 PS Claim 4; Fig 4C; 110pp; English.

XX
 CC The present sequence encodes a consensus human bikunin, which

XX
 CC inhibits, e.g., trypsin, Kallikrein, plasmin and factor Xlla.

XX
 CC Bikunin can be used to treat or prevent brain and spinal cord

XX
 CC oedema, inflammation, infection or granulomatosis, multiple

XX
 CC sclerosis, ischaemia, perioperative blood loss, sepsis, shock,

XX
 CC fibrosis, blood coagulation diseases, polytrauma, stroke,

XX
 CC cerebral or subarachnoid haemorrhage and gastric or cervical

XX
 CC cancer and prevent metastasis. It is particularly useful for

XX
 CC reducing blood loss during surgery, and can also be used to treat

XX
 CC other cancer, arthritis, anaemia, non-insulin dependent diabetes,

XX
 CC influenza and similar viral infections, acute pancreatitis and

XX
 CC gout, and prevent pre-term labour. It has similar properties to

XX
 CC aprotinin, but is less highly charged so should be less

XX
 CC immunogenic and less likely to damage the kidneys. Manipulation

XX
 CC of the bikunin sequence may allow the inhibitory profile to be

XX
 CC altered. It also reduces or eliminates the need for whole donor

XX
 CC blood or blood products during surgery, thereby reducing the risk

XX
 CC of infection and other adverse side effects, as well as reducing

XX
 CC the cost of surgery.

XX
 SQ Sequence 1528 BP; 320 A; 386 C; 467 G; 352 T; 3 other;

Query Match 92.6%; Score 808; DB 18; Length 1528;
 Best Local Similarity 97.5%; Pred. No. 9,7e-194;
 Matches 842; Conservative 0; Mismatches 20; Indels 2; Gaps 2;

QY 1 CTCGAGCATATGTTCACTATGAGAAATACAGCCGCAAGCAGCTACCTGGGCTTG 60
 DB 666 CTCGAGCATATGTTCACTATGAGAAATACAGCCGCAAGCAGCTACCTGGGCTTG 725
 QY 61 CCGTGACATCTTCCACGCTGTGATCTTGAAGTGAAGAGAACTCCCTGCAATTAATTCT 120
 DB 726 CCGTGACATCTTCCACGCTGTGATCTTGAAGTGAAGAGAACTCCCTGCAATTAATTCT 785
 QY 121 CTATGAGGCTGCGGAGGCAATTAAGAAAGCTACCTGAGAGGCTGCAATGCTCCG 180
 DB 786 CTATGAGGCTGCGGAGGCAATTAAGAAAGCTACCTGAGAGGCTGCAATGCTCCG 845
 QY 181 CTGCTCCGAGAGAGAAATCTCCCTGCTGCTGAGAGCTGCAATGAGTGTCTGCG 240
 DB 846 CTGCTCCGAGAGAGAAATCTCCCTGCTGCTGAGAGCTGCAATGAGTGTCTGCG 905
 QY 241 GGGGCTGTCTGATGATGATCTCTCTGAGAGCTCCATGATGATGATGATGATGAT 300
 DB 906 GGGGCTGTCTGATGATGATCTCTCTGAGAGCTCCATGATGATGATGATGATGAT 965
 QY 301 GGTGACAGGAGAAACAGAGAGCTGCTGCGCACTGCTGAGAGCTGCGAGATGCAA 360
 DB 966 GGTGACAGGAGAAACAGAGAGCTGCTGCGCACTGCTGAGAGCTGCGAGATGCAA 1025
 QY 361 GGAGCAGCTGCTGAGAAACATATGCTGCTGAGAGCTGCTGCGCAAGAGACT-866 419
 DB 1026 GGAGCAGCTGCTGAGAAACATATGCTGCTGAGAGCTGCTGCGCAAGAGACT-866 1085
 QY 420 AAGGAGAGGAGAACTATGCTGAGCTTTTAAATGAAGGAGATTGACTCGGATTTAGT 479
 DB 1086 AAGGAGAGGAGAACTATGCTGAGCTTTTAAATGAAGGAGATTGACTCGGATTTAGT 1145
 QY 480 GATCATTAAGGCTGAGCTGCTTTCTCTGAGAGCTGAGAGCTGCTTCTGCTGCGCA 539
 DB 1146 GATCATTAAGGCTGAGCTGCTTTCTCTGAGAGCTGAGAGCTGCTTCTGCTGCGCA 1205
 QY 540 GGGATGAGGCTGCTGAGAAATCTCTGAGAGGCTCTCTGCTGAGAGCTGCTGAGT 599
 DB 1206 GGGATGAGGCTGCTGAGAAATCTCTGAGAGGCTCTCTGCTGAGAGCTGCTGAGT 1265
 QY 600 CAGCAGCCCGAGTGTCTCTGCTGAGTCAATTTCTTCTGAGAGTGTCTTCTT 659
 DB 1266 CAGCAGCCCGAGTGTCTCTGCTGAGTCAATTTCTTCTGAGAGTGTCTTCTT 1325
 QY 660 GCTTATGTTGAATTCATTCCTCTTCTGATCAGAGAGTATGAGATGCTTCT 719
 DB 1326 GCTTATGTTGAATTCATTCCTCTTCTGATCAGAGAGTATGAGATGCTTCT 1384
 QY 720 TTTGTTGCTGATTTAGTTTCTTAAATGAAGTCTTCTTAAAGATTTTAAAGATTTCTG 779
 DB 1385 TTTGTTGCTGATTTAGTTTCTTAAATGAAGTCTTCTTAAAGATTTTAAAGATTTCTG 1444
 QY 780 AAGAGAGAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 839
 DB 1445 AAGAGAGAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1504
 QY 840 TTCAGCATGCTCTTCAAAAAA 863
 DB 1505 AAAAAAATCAAAAAA 1528

RESULT 13
 ID ABA83112
 AC ABA83112;
 XX ABA83112;
 DT 08-FEB-2002 (first entry)
 XX

DE Kunitz type 2 Ser protease inhibitor ovarian tumour marker gene, #61.
 XX
 XX Ovarian tumour marker gene; human; overexpression; upregulation;
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 KW identification; serous cystadenoma; borderline serous tumour;
 KW mucinous cystadenocarcinoma; mucinous mucinous tumour; endometrioid carcinoma;
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
 KW immune response pathway; cell proliferation regulation; protein folding;
 KW membrane localised; secreted; therapeutic target; cytostatic;
 KW gene therapy; vaccine; ds.
 XX
 OS Homo sapiens.
 PN WO200175177-A2.
 PD 11-OCT-2001.
 XX
 XX 03-APR-2001; 2001WO-US10947.
 PF 03-APR-2000; 2000US-194336P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Morin PJ, Sherman-Baust CA, Pizer SS, Hough CD,
 PI WPI, 2001-626450/72.
 DR P-PSDB; ABB50286.
 XX
 XX Detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian
 PT cancer treatment, by measuring expression level of ovarian tumor marker
 PT gene -
 PS Claim 23; Page 104-105; 140pp; English.
 XX
 XX The invention relates to methods for diagnosing and prognosing ovarian
 CC tumors in an individual via the detection and measurement of the
 CC expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
 CC ABA83181 and ABA83183). The methods of the invention are useful for
 CC detecting an ovarian tumor in a patient, for identifying an individual
 CC at increased risk for developing ovarian cancer, in tests for
 CC assessing the relative severity of ovarian cancer, in tests for
 CC monitoring disease status in a patient being treated for ovarian cancer,
 CC monitoring the disease status in a patient being treated for ovarian cancer.
 CC The methods can additionally be used to identify a particular tumor as
 CC being an ovarian tumor (i.e., an epithelial ovarian tumor selected from
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
 CC mucinous cystadenoma, endometrioid carcinoma, undifferentiated carcinoma,
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
 CC tumour). The ovarian tumor marker genes of the invention were identified
 CC using SAGE (serial analysis of gene expression) and were found to be
 CC overexpressed in a broad variety of ovarian epithelial tumour cells
 CC relative to normal ovarian epithelial cells. The marker genes are
 CC implicated in immune response pathways, in the regulation of cell
 CC proliferation and in protein folding, and many of these are membrane-
 CC localised or secreted. In addition to their use as diagnostic and
 CC prognostic markers, the ovarian tumour marker genes or their encoded
 CC proteins may be used as therapeutic targets for the treatment and
 CC prevention of ovarian cancer. Sequences ABA83081-ABA83122, ABA83180,
 CC ABA83182 and ABA83184 represent the ovarian tumour marker genes of
 CC the invention.
 CC
 XX
 XX Sequence 1564 BP; 297 A; 419 C; 491 G; 357 T; 0 other;
 SQ

Query Match 92.3%; Score 805.6; DB 22; Length 1564;
 Best Local Similarity 99.3%; Pred. No. 3.9e-133;
 Matches 830; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 CTCGAGCATATGTTCACTATGAGAAATACAGCCGCAAGCAGCTACCTGGGCTTG 60

Db 730 CTCACGATATGTTCAATATGAAATCTGACACCGCAACGAGTCACTGGGCTTG 789
 QY 61 CCGTGATCTCTCCACGCTGTGACTTGAAGTGAAGAACTCTGCAATACTTCAT 120
 Db 790 CCGTGATCTCTCCACGCTGTGACTTGAAGTGAAGAACTCTGCAATACTTCAT 849
 QY 121 CTATGAGGCTCCCGGGCAATTAAGAACACTACCGCTCTGAGAGGCTCTGATGTCG 180
 Db 850 CTATGAGGCTCCCGGGCAATTAAGAACACTACCGCTCTGAGAGGCTCTGATGTCG 909
 QY 181 CTGCTTCGCCGACGAGAGAACTCTCCCGCCCTTGGCTCAAGGAGTGGTGTG 240
 Db 910 CTGCTTCGCCGACGAGAGAACTCTCCCGCCCTTGGCTCAAGGAGTGGTGTG 969
 QY 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 Db 970 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1029
 QY 301 GGTGACACGAGAGAACTCAAGAGCTGCTGACACCTCTGAGCTTCGAGATGACAA 360
 Db 1030 GGTGACACGAGAGAACTCAAGAGCTGCTGACACCTCTGAGCTTCGAGATGACAA 1089
 QY 361 GAGACAGCTGTGAGAGAACTCAAGAGCTGCTGACACCTCTGAGCTTCGAGATGACAA 419
 Db 1090 GAGACAGCTGTGAGAGAACTCAAGAGCTGCTGACACCTCTGAGCTTCGAGATGACAA 1149
 QY 420 AAGGAGAGGAGAGAACTCAAGAGCTGCTGACACCTCTGAGCTTCGAGATGACAA 479
 Db 1150 AAGGAGAGGAGAGAACTCAAGAGCTGCTGACACCTCTGAGCTTCGAGATGACAA 1209
 QY 480 GATCATTAGGCTGAGAGCTGCTGCTGAGAGAGTGAAGAGCTGCTGCTGCTGCTG 539
 Db 1210 GATCATTAGGCTGAGAGCTGCTGCTGAGAGAGTGAAGAGCTGCTGCTGCTGCTG 1269
 QY 540 GGGATGGGTTCTGTAATCTCTGAGAGGCTCTCTCTGAGAGGCTGAGAGCTGCTG 599
 Db 1270 GGGATGGGTTCTGTAATCTCTGAGAGGCTCTCTCTGAGAGGCTGAGAGCTGCTG 1329
 QY 600 CAGACAGGCTGAGAGCTGCTCTCTGAGAGGCTCTCTCTGAGAGGCTGAGAGCTG 659
 Db 1330 CAGACAGGCTGAGAGCTGCTCTCTGAGAGGCTCTCTCTGAGAGGCTGAGAGCTG 1389
 QY 660 GCTTATGTAATTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719
 Db 1390 GCTTATGTAATTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1449
 QY 720 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 779
 Db 1450 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1508
 QY 780 AAAG 835
 Db 1509 AAAG 1564

RESULT 14
 AACT78030
 ID AACT78030 standard; CDNA; 1649 BP.
 AC AACT78030;
 XX
 XX 08-FEB-2001 (first entry)
 DE Human cancer associated gene sequence SEQ ID NO:424.
 XX
 XX Human cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
 KW antidiabetic; antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW antiinflammatory; antihypertensive; antidiabetic; antidiabetic; antidiabetic;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KW vasotropic; antiproliferative; antiproliferative; antiproliferative;
 KW immune disorder; hematopoietic cell disorder; autoimmune disorder;

KM allergic reaction; graft versus host disease; organ rejection;
 KM hemostatic; thrombolytic; cardiovascular disorder; infection;
 KM neurological disease; drug screening; ss.
 OS Homo sapiens.
 XX
 XX W020005350-A1.
 PN
 XX 21-SEP-2000.
 PD
 XX 08-MAR-2000; 2000WO-US05882.
 PF
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX WPI: 2000-587533/55.
 DR
 XX P-PSDB; AAB43821.
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 useful for treating or diagnosing e.g. cancer -
 PS
 XX Claim 1; Page 965; 2352pp; English.
 CC AACT7607 to AACT78449 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnerability; immunomodulator;
 CC antidiabetic; antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 CC antineoplastic; antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 CC neuroprotective; cardioprotective; thrombolytic; coagulant;
 CC dermatological; vasotropic; antiproliferative; antiproliferative; antiproliferative;
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AACT78449 to
 CC AACT7457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.
 CC
 XX
 XX Sequence 1649 BP; 335 A; 443 C; 501 G; 369 T; 1 other;
 Query Match 86.8%; Score 758.2; DB 21; Length 1649;
 Best Local Similarity 95.6%; Pred. No. 3.7e-181;
 Matches 833; Conservative 0; Mismatches 33; Indels 5; Gaps 5;
 QY 1 CTCCAGCATATGTTCAATATGAAATCTGACACCGCAACGAGTCACTGGGCTTG 60
 Db 784 CTCACGATATGTTCAATATGAAATCTGACACCGCAACGAGTCACTGGGCTTG 843
 QY 61 CCGTGATCTCTCCACGCTGTGACTTGAAGTGAAGAACTCTGCAATACTTCAT 120
 Db 844 CCGTGATCTCTCCACGCTGTGACTTGAAGTGAAGAACTCTGCAATACTTCAT 903
 QY 121 CTATGAGGCTCCCGGGCAATTAAGAACACTACCGCTCTGAGAGGCTCTGATGTCG 180
 Db 904 CTATGAGGCTCCCGGGCAATTAAGAACACTACCGCTCTGAGAGGCTCTGATGTCG 963
 QY 181 CTGCTTCGCCGACGAGAGAACTCTCCCGCCCTTGGCTCAAGGAGTGGTGTG 240
 Db 964 CTGCTTCGCCGACGAGAGAACTCTCCCGCCCTTGGCTCAAGGAGTGGTGTG 1023
 QY 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 Db 1024 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1083

```

OY 301 GGTGGACGAGGAAACGAGACCGTGCCTGCGACCTCTGAGCTCCGAGATGACAA 360
DB 1084 GGTGGACGAGGAAACGAGACCGTGCCTGCGACCTCTGAGCTCCGAGATGACAA 1143
OY 361 GAGACAGCTGCTAGAACACATATGCTCTGTGACCGCCCTGTGCGCAAGAGATGGGA 420
DB 1144 GAGACAGCTGCTAGAACACATATGCTCTGTGACCGCCCTGTGCGCAAGAGATGGGA 1202
OY 421 AGGAGGGGAGACTATGCTGAGCTTTTAAATAGAGGATTAACCTGCAATTTAGAG 480
DB 1203 AAGGAGGGGAGACTATGCTGAGCTTTTAAATAGAGGATTAACCTGCAATTTAGAG 1261
OY 481 ATCATTAAGGCTGAGCTGTTCTCTGAGAGGTAGACGAGCTGCTCTGCTGCTGAG 540
DB 1262 ATCATTAAGGCTGAGCTGTTCTCTGAGAGGTAGACGAGCTGCTCTGCTGCTGAG 1321
OY 541 GGAATGGCTTGTCTTGGAAATCTCTAGAGAGGCTCTCTGCTGCTGCTGAGCTGAG 600
DB 1322 GGAATGGCTTGTCTTGGAAATCTCTAGAGAGGCTCTCTGCTGCTGCTGAGCTGAG 1381
OY 601 AGCAGCCCGGAGTGTCTCTGAGATGATTTCTTCTCTGAGTGAATTTCTTTG 660
DB 1382 AGCAGCCCGGAGTGTCTCTGAGATGATTTCTTCTCTGAGTGAATTTCTTTG 1440
OY 661 CTATATGTTGAATTCATTCCTCTTCTCTATACAGAAATGATGTTGAAATGCTTCT 720
DB 1441 CTATATGTTGAATTCATTCCTCTTCTCTATACAGAAATGATGTTGAAATGCTTCT 1499
OY 721 TTGTTTGTCTGATTTATGTTTATTAAGTATTAACAAAGTTTATTAAGCTTCTGA 780
DB 1500 TTGTTTGTCTGATTTATGTTTATTAAGTATTAACAAAGTTTATTAAGCTTCTGA 1559
OY 781 AAGAGAGAAATGTTTATTAAGTATTAACAAAGTTTATTAAGCTTCTGAATTAAT 840
DB 1560 AAGAGAGAAATGTTTATTAAGTATTAACAAAGTTTATTAAGCTTCTGAATTAAT 1618
OY 841 TCAGCATGTCTTCAAAAAA 871
DB 1619 AAAAAA 1649

RESULT 15
AA16487
ID AA16487 standard; DNA; 597 BP.
XX
AC AA16487;
XX
DT 14-JUN-2000 (first entry)
XX
DE Human colon cancer differentially expressed nucleotide sequence #492.
XX
KW Colon cancer; detect; differential expression; human; treatment;
XX
KW detect mutation; non-invasive diagnostic method; ds.
XX
OS Homo sapiens.
XX
PN WO200012702-A2.
XX
PD 09-MAR-2000.
XX
PF 30-AUG-1999; 99WO-US19424.
XX
PR 31-AUG-1998; 98US-0098639.
XX
PR 27-JAN-1999; 99US-0117393.
XX
PA (FAR3) BAYER CORP.
XX
PI Endege NO, Steimann KE, Astle JH, Burgess CC, Carroll E,
PI Catino T, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JB,
PI Schlegel R;
XX
DR WPI, 2000-256641/22.

```

```

XX
PT Novel nucleic acids and proteins for identifying therapeutic agents
XX
XX useful for treating and diagnosing cancer, especially colon cancer -
XX
PS Claim 16; Page 323; 345pp; English.
XX
CC This sequence represents a human nucleotide sequence which is
CC differentially expressed in colon cancer cells compared to the expression
CC levels in normal cells. The nucleotide sequence can be used as a source
CC of primers and probes. The nucleotide sequence is useful for determining
CC the phenotype of a cell by detecting the differential expression of the
CC sequence relative to a normal cell. The probes derived from the sequence
CC can also be used to determine the phenotype of cells in a sample. Probes
CC and antibodies which hybridize to the nucleotide sequence can also be
CC used to determine the phenotype of a cell. The primers are useful for
CC detecting a mutation in a test nucleotide sequence and also for detecting
CC cancer, preferably colon cancer. Antibodies against the protein encoded
CC by the nucleotide sequence can also be used in a method to detect colon
CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
CC colon cancer at an early stage.
XX
XX
SO Sequence 597 BP; 121 A; 117 C; 157 G; 162 T; 40 other;
XX
XX
Query Match 48.7%; Score 425.4; DB 21; Length 597;
Best Local Similarity 93.3%; Pred. No. 2.e-97;
Matches 460; Conservative 0; Mismatches 31; Indels 2; Gaps 2;
OY 300 GGTGGACGAGGAAACGAGACCGTGCCTGCGACCTCTGAGCTCCGAGATGACAA 359
DB 8 GGTGGACGAGGAAACGAGACCGTGCCTGCGACCTCTGAGCTCCGAGATGACAA 67
OY 360 AGAGAGAGCTGCTAGAACACATATGCTCTGACCGCCCTGTGCGCAAGAGACT-GG 418
DB 68 AGAGAGAGCTGCTAGAACACATATGCTCTGACCGCCCTGTGCGCAAGAGACTGG 127
OY 419 GAAGGAGAGGAGACTATGCTGAGCTTTTAAATAGAGGATTAACCTGCAATTTAG 478
DB 128 GAAGGAGAGGAGACTATGCTGAGCTTTTAAATAGAGGATTAACCTGCAATTTAG 187
OY 479 TGATCATTAAGGCTGAGCTGTTTCTCTGAGAGGTAGACGAGCTGCTCTGAGCTGGC 538
DB 188 TGATCATTAAGGCTGAGCTGTTTCTCTGAGAGGTAGACGAGCTGCTCTGAGCTGGC 247
OY 539 AGGAGATGGCTTGTCTTGGAAATCTCTAGAGAGCTCTCTGCAATGCTGCACTG 598
DB 248 AGGAGATGGCTTGTCTTGGAAATCTCTAGAGAGCTCTCTGCAATGCTGCACTG 307
OY 599 GCAGCAGCCCGGAGTGTCTCTGAGAGGTAGACGAGCTCTCTGCAATGCTGCACTG 658
DB 308 GCAGCAGCCCGGAGTGTCTCTGAGAGGTAGACGAGCTCTCTGCAATGCTGCACTG 367
OY 659 TGCTTATGTTGAATTCATTCCTCTTCTCTATACAGAAATGATGTTGAAATGCT 717
DB 368 TGCTTATGTTGAATTCATTCCTCTTCTCTATACAGAAATGATGTTGAAATGCT 427
OY 718 CTTTGTGCTGATTTATGTTTATTAAGTATTAACAAAGTTTATTAAGCTTCTGA 777
DB 428 CTTTGTGCTGATTTATGTTTATTAAGTATTAACAAAGTTTATTAAGCTTCTGA 487
OY 778 TGAAGAAGGAAA 790
DB 488 TTAANAAGGAAA 500

```

Search completed: January 10, 2003, 01:53:34
Job time: 261.933 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:47:01 (Search time 51.01 Seconds

(without alignments)
5248.554 Million cell updates/sec

Title: US-09-825-682A-56

Perfect score: 873

Sequence: 1 ctccagcagatcgttcaact.....tcaaaaaaaaaaaaaaaaaa 873

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A_5B.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/5A_5B.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/5A_5B.COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862	98.7	1610	4	US-09-013-896A-1
2	425.4	48.7	597	4	US-09-385-982-492
3	394	45.1	759	1	US-08-685-660A-4
4	394	45.1	759	2	US-08-974-196-4
5	282.4	32.3	287	4	US-09-013-896A-13
6	260	29.8	273	4	US-09-013-896A-14
7	250	28.6	256	4	US-09-013-896A-15
8	188	21.5	201	4	US-09-013-896A-17
9	167.4	19.2	276	4	US-09-404-879A-169
10	167	19.1	207	4	US-09-404-879A-356
11	167	19.1	371	4	US-09-404-879A-365
12	165.4	18.9	276	4	US-09-404-879A-168
13	146.8	16.8	199	4	US-09-222-575-125
14	146.8	16.8	783	4	US-09-020-956-15
15	146.8	16.8	783	4	US-09-030-607-15
16	146.8	16.8	783	4	US-09-605-785-15
17	146.8	16.8	783	4	US-09-439-313-15
18	146.8	16.8	783	4	US-09-352-616A-15
19	146.8	16.8	783	4	US-09-232-149A-15
20	146.8	16.8	783	4	US-09-232-149A-15
21	146.8	16.8	783	4	US-09-232-149A-15
22	146.8	16.8	783	4	US-09-232-149A-15
23	146.8	16.8	783	4	US-09-232-149A-15
24	146.8	16.8	783	4	US-09-232-149A-15
25	146.8	16.8	783	4	US-09-232-149A-15
26	146.8	16.8	783	4	US-09-232-149A-15
27	146.8	16.8	783	4	US-09-232-149A-15

28	56.6	6.5	245	6	5223482-28	Patent No. 5223482
29	55.8	6.4	704	2	US-08-829-876-100	Sequence 100, App
30	55.8	6.4	704	2	US-08-829-876-104	Sequence 104, App
31	55.8	6.4	704	2	US-09-234-874A-100	Sequence 100, App
32	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
33	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
34	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
35	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
36	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
37	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
38	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
39	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
40	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
41	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
42	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
43	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
44	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App
45	55.8	6.4	704	2	US-09-234-874A-104	Sequence 104, App

ALIGNMENTS

RESULT 1
US-09-013-896A-1
Sequence 1, Appl
Patent No. 5223482
GENERAL INFORMATION:
APPLICANT: GENZ, REINER
TITLE OF INVENTION: TISSUE FACTOR PATHWAY INHIBITOR-3
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., NW, STB. 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/013,896A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1290001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1610 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 361..1116
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 361..439
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 442..1116
US-09-013-896A-1
Query Match 98.7%, Score 862, DB 4, Length 1610;

Best Local Similarity 99.9%; Pred. No. 1e-211; Matches 873; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 CTCACGATATGTTCACTAATGAAATCTGCAACCGCAACGAGCTCGGCGCTTG 60
Db 726 CTCACGATATGTTCACTAATGAAATCTGCAACCGCAACGAGCTCGGCGCTTG 785
QY 61 CCGTGATCTCTCCACCGCTGTAATTTGAAGTGAAGAACTCTGCAATTAATTCAT 120
Db 786 CCGTGATCTCTCCACCGCTGTAATTTGAAGTGAAGAACTCTGCAATTAATTCAT 845
QY 121 CTATGAGAGCTCCCGGCAATTAAGAACACTACCGCTCTGAGAGCTGATGCTCG 180
Db 846 CTATGAGAGCTCCCGGCAATTAAGAACACTACCGCTCTGAGAGCTGATGCTCG 905
QY 181 CTGCTTCCCGCAGAGAGAAATCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db 906 CTGCTTCCCGCAGAGAGAAATCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 965
QY 241 GGGGCTGTTGCTGATGCTGATCTCTCTGAGAGCTGATGCTGATGCTGATGCTG 300
Db 966 GGGGCTGTTGCTGATGCTGATCTCTCTGAGAGCTGATGCTGATGCTGATGCTG 1025
QY 301 GGTGGACGAGAGAACAGAGAGCTGCTGCGACCTGCTGAGAGCTCGAGATGACAA 360
Db 1026 GGTGGACGAGAGAACAGAGAGCTGCTGCGACCTGCTGAGAGCTCGAGATGACAA 1085
QY 361 GAGACAGCTGCTGAAAGAACATATGCTCTGAGACCGCCCTTCCGCAAGAGACT- 419
Db 1086 GAGACAGCTGCTGAAAGAACATATGCTCTGAGACCGCCCTTCCGCAAGAGACT- 1145
QY 420 AAGGAGAGGAGACTATGAGCTTTTAAATAGAGGATGATCGGATTTGAGT 479
Db 1146 AAGGAGAGGAGACTATGAGCTTTTAAATAGAGGATGATCGGATTTGAGT 1205
QY 480 GATCATTAGAGGCTGAGGCTGCTTCTGAGAGGATGAGAGAGCTCTTCTGCTGCGA 539
Db 1206 GATCATTAGAGGCTGAGGCTGCTTCTGAGAGGATGAGAGAGCTCTTCTGCTGCGA 1265
QY 540 GGGATGGGTTGCTTGAAGAAATCTCTGAGAGGCTCTCTGAGATGGCTGAGTCTG 599
Db 1266 GGGATGGGTTGCTTGAAGAAATCTCTGAGAGGCTCTCTGAGATGGCTGAGTCTG 1325
QY 600 CAGAGCCCGAGAGTGTCTCTGCTGATGATTTCTTCCGAGAGTGAATTTCTT 659
Db 1326 CAGAGCCCGAGAGTGTCTCTGCTGATGATTTCTTCCGAGAGTGAATTTCTT 1385
QY 660 GCTTATGTTGAATTCATTCGCTCTTCTCTCATCAGAGAGTATGTTGAATGCTTCT 719
Db 1386 GCTTATGTTGAATTCATTCGCTCTTCTCTCATCAGAGAGTATGTTGAATGCTTCT 1445
QY 720 TTTGTTGCTGATTTATGTTTAAAGTAAACAAAGTTTAAATAGATCTG 779
Db 1446 TTTGTTGCTGATTTATGTTTAAAGTAAACAAAGTTTAAATAGATCTG 1505
QY 780 AAAAGAGAGAAATGTAAGATTAATTAAGAGAGGCTTCCCTCTAGAAATAAT 839
Db 1506 AAAAGAGAGAAATGTAAGATTAATTAAGAGAGGCTTCCCTCTAGAAATAAT 1565
QY 840 TTCAGCATGTGCTTTCAAAAAATTTTTAAAAA 873
Db 1566 TTCAGCATGTGCTTTCAAAAAATTTTTAAAAA 1599

```

RESULT 2
US-09-385-982-492
; Sequence 492, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: EMBERGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCNNA-260XX

;; CURRENT APPLICATION NUMBER: US/09/385,982
;; CURRENT FILING DATE: 1999-08-30
;; EARLIER APPLICATION NUMBER: 09/328,111
;; EARLIER FILING DATE: 1999-06-08
;; EARLIER APPLICATION NUMBER: 60/117,393
;; EARLIER FILING DATE: 1998-01-27
;; EARLIER APPLICATION NUMBER: 60/098,639
;; EARLIER FILING DATE: 1998-08-31
;; NUMBER OF SEQ ID NOS: 544
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 492
;; LENGTH: 597
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURES:
;; NAME/KEY: misc.feature
;; LOCATION: (1) -- (597)
;; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-492

Query Match 48.7%; Score 425.4; DB 4; Length 597;
Best Local Similarity 93.3%; Pred. No. 5.6e-100;
Matches 460; Conservative 0; Mismatches 31; Indels 2; Gaps 2;

```

QY 300 GGGTGGACGAGAGAACAGAGAGCTGCTGCGACCGCTCTGAGAGCTCGGAGATGACA 359
Db 8 GGGTGGACGAGAGAACAGAGAGCTGCTGCGACCGCTCTGAGAGCTCGGAGATGACA 67
QY 360 AGAGAGAGCTGAGAGAACACATATGCTCTGAGACCGCCCTGCGCAAGAGACT- 418
Db 68 AGAGAGAGCTGAGAGAACACATATGCTCTGAGACCGCCCTGCGCAAGAGACT- 127
QY 419 GAAAGAGAGGAGACTATGAGCTTTTAAATAGAGGATGATCGGATTTGAG 478
Db 128 GAAAGAGAGGAGACTATGAGCTTTTAAATAGAGGATGATCGGATTTGAG 187
QY 479 TGATCATTAAGGCTGAGCTGCTTCTCTGAGAGTGAAGAGCTCTTCTGCTGCTG 538
Db 188 TGATCATTAAGGCTGAGCTGCTTCTCTGAGAGTGAAGAGCTCTTCTGCTGCTG 247
QY 539 AGGATGGGTTGCTTGAAGAAATCTCTAGAGAGCTCTCTCGATGAGCTGAGCTG 598
Db 248 AGGATGGGTTGCTTGAAGAAATCTCTAGAGAGCTCTCTCGATGAGCTGAGCTG 307
QY 599 GCACAGACCCCGAGTGTCTCTGCTGATGATTTCTTCCAGAGTGAATTTCTT 658
Db 308 GCACAGACCCCGAGTGTCTCTGCTGATGATTTCTTCCAGAGTGAATTTCTT 367
QY 659 TGCTTATGTTGAATTCATTCGCTCTTCTCTCATCAGAGAGTATGTT- 717
Db 368 TGCTTATGTTGAATTCATTCGCTCTTCTCTCATCAGAGAGTATGTT- 427
QY 718 CTTTGTTGCTGATTTATGTTTAAAGTAAACAAAGTTTAAATAGATCTG 777
Db 428 CTTTGTTGCTGATTTATGTTTAAAGTAAACAAAGTTTAAATAGATCTG 487
QY 778 TGAAGAAGGAAA 790
Db 488 TTAANNAAGGAA 500

```

RESULT 3
US-08-685-660A-4
; Sequence 4, Application US/0868560A
; Patent No. 5731412
; GENERAL INFORMATION:
; APPLICANT: SHIMOWURA, Takeshi
; APPLICANT: KAWAGUCHI, Toshiya
; APPLICANT: KITAMURA, Naomi
; TITLE OF INVENTION: NOVEL PROTEIN, DNA CODING FOR SAME
; TITLE OF INVENTION: AND METHOD OF PRODUCING THE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

```

ADDRESSER: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,660A
FILING DATE: 24-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JPA Hei 7-187134
FILING DATE: 24-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: Q-42295
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7860
TELEFAX: (202) 293-7060
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: MKN45
FEATURE:
NAME/KEY: coding sequence
LOCATION: 1 to 759
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal peptide
LOCATION: 1 to 81
IDENTIFICATION METHOD: by experiment
NAME/KEY: mature peptide
LOCATION: 82 to 759
IDENTIFICATION METHOD: by experiment
US-08-685-660A-4

Query Match          45.1%; Score 394; DB 1; Length 759;
Best Local Similarity 100.0%; Pred. No. 7.1e-92;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CTCAGGATATGTTCAACTATGAAGAAATCTGACCGCAACGAGTCACTG394CTTG 60
 DB 366 CTCAGGATATGTTCAACTATGAAGAAATCTGACCGCAACGAGTCACTG394CTTG 425
 QY 61 CCGTGCACTCTTCCACGCTGTCTTGAAGTGAGAGAACTCTGCAATACTTCA 120
 DB 426 CCGTGCACTCTTCCACGCTGTCTTGAAGTGAGAGAACTCTGCAATACTTCA 485
 QY 121 CTATGAGAGCTGCGG393CAATAAAGACCTACCGCTCTGAGAG393CTTGCTCG 180
 DB 486 CTATGAGAGCTGCGG393CAATAAAGACCTACCGCTCTGAGAG393CTTGCTCG 545
 QY 181 CTGCTTCCGACAGAGAAATCTCCCTGCTGCTGCTCAAGGTGTGTCTTGCG 240
 DB 546 CTGCTTCCGACAGAGAAATCTCCCTGCTGCTGCTCAAGGTGTGTCTTGCG 605
 QY 241 GGGGCTTTTGTGATGTGTGATCTTCTCTG394AGCTTCCATGCTCTCATGCTCG 300
 DB 606 GGGGCTTTTGTGATGTGTGATCTTCTCTG394AGCTTCCATGCTCTCATGCTCG 665
 QY 301 GGTGGACGAGAGAAACAGAGAGCTGCGCGCACCGCTCTGAGCTCGGAGATGACA 360

```

DB 666 GGTGGACGAGAGAAACAGAGAGCTGCGCGCACCGCTCTGAGCTCGGAGATGACA 725
QY 361 GGAGCAGCTGTGTAAGAAACATATGCTCTGTA 394
DB 726 GGAGCAGCTGTGTAAGAAACATATGCTCTGTA 759

RESULT 4
US-08-974-196-4
Sequence 4, Application US/08974196
Patent No. 5854396
GENERAL INFORMATION:
APPLICANT: SHIMOMURA, Takeshi
APPLICANT: KAWAGUCHI, Toshiya
APPLICANT: KITAMURA, Naomi
TITLE OF INVENTION: NOVEL PROTEIN, DNA CODING FOR SAME
TITLE OF INVENTION: AND METHOD OF PRODUCING THE PROTEIN
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,196
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/685,660
FILING DATE: 24-JUL-1996
APPLICATION NUMBER: JPA Hei 7-187134
FILING DATE: 24-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: Q-42295
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7860
TELEFAX: (202) 293-7060
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: MKN45
FEATURE:
NAME/KEY: coding sequence
LOCATION: 1 to 759
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal peptide
LOCATION: 1 to 81
IDENTIFICATION METHOD: by experiment
NAME/KEY: mature peptide
LOCATION: 82 to 759
IDENTIFICATION METHOD: by experiment
US-08-974-196-4

Query Match          45.1%; Score 394; DB 2; Length 759;
Best Local Similarity 100.0%; Pred. No. 7.1e-92;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CTCAGGATATGTTCAACTATGAAGAAATCTGACCGCAACGAGTCACTG394CTTG 60

Db 366 CTCACGATATGTCTCACTATGAGATATGACCGCCGACAGCATCTACGCGCTTG 425
Qy 61 CCGTACCTCTTCCACGCTGTACTTTGACGTGAGAGAACTCTCTCATTAATCTCAT 120
Db 426 CCGTACCTCTTCCACGCTGTACTTTGACGTGAGAGAACTCTCTCATTAATCTCAT 485
Qy 121 CTATGAGAGCTCCGGGGCATTAAGACATACCTCTTGAGAGAGCTCTGATGCTCCG 180
Db 486 CTATGAGAGCTCCGGGGCATTAAGACATACCTCTTGAGAGAGCTCTGATGCTCCG 545
Qy 181 CTGCTTCGACAGAGAGATCTCTCCCTGCTCCCTGCTGCTCAAGAGTGTGCTTGAC 240
Db 546 CTGCTTCGACAGAGAGATCTCTCCCTGCTCCCTGCTGCTCAAGAGTGTGCTTGAC 605
Qy 241 GGGGCTGTCTGTATGATGTATCTCTCTTGAGAGCTCTCATGCTGATCTCG 300
Db 606 GGGGCTGTCTGTATGATGTATCTCTCTTGAGAGCTCTCATGCTGATCTCG 665
Qy 301 GGTGACAGAGAGAACAGAGAGCTGCTGACCTGCTGAGCTCCGAGATGACAA 360
Db 666 GGTGACAGAGAGAACAGAGAGCTGCTGACCTGCTGAGCTCCGAGATGACAA 725
Qy 361 GAGAGAGCTGTGTAAGACATATGCTCTGTA 394
Db 726 GAGAGAGCTGTGTAAGACATATGCTCTGTA 759

RESULT 5
US-09-013-896A-13

; Sequence 13, Application US/09013896A
; Patent No. 6262233
; GENERAL INFORMATION:
; APPLICANT: GENTZ, REINER
; TITLE OF INVENTION: TISSUE FACTOR PATHWAY INHIBITOR-3
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, STE. 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,896A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-013-896A-13

Query Match 32.3%; Score 282.4; DB 4; Length 287;
Best Local Similarity 98.6%; Pred. No. 1,9e-63;
Matches 283; Conservative 0; Mismatches 4; Indels 0;
Gaps 0;
Qy 46 ATCTACCTGGGCTTGGCCGTCATCTTCCACGCTGTACTTTGACGTGAGAGAACTC 105

Db 1 AGTACCTGGGCTTGGCCGTCATCTTCCACGCTGTACTTTGACGTGAGAGAACTC 60
Qy 106 CTGCAATAATCTTATCTATGAGAGCTGCGGGCAATAAAGAACATCACTCTGAGAG 165
Db 61 CTGCAATAATCTTATCTATGAGAGCTGCGGGCAATAAAGAACATCACTCTGAGAG 120
Qy 166 GGCCTGATCTCTCGCTGCTTCCGACAGAGAAATCTCTCCCTGCCCCCTTGCTCAAA 225
Db 121 GGCCTGATCTCTCGCTGCTTCCGACAGAGAAATCTCTCCCTGCCCCCTTGCTCAAA 180
Qy 226 GGTGATGCTTCTGCGGGGCTGTTCGTATGATGTGATCTCTCTCTGAGAGCTTCAT 285
Db 181 GATGATGCTTCTGCGGGGCTGTTCGTATGATGTGATCTCTCTCTGAGAGCTTCAT 240
Qy 286 GGTCTACCTGATCCGGGTGACAGAGAAACAGAGAGCTGCTGCTC 332
Db 241 GGTCTACCTGATCCGGGTGACAGAGAAACAGAGAGCTGCTGCTC 287

RESULT 6

US-09-013-896A-14/c

; Sequence 14, Application US/09013896A
; Patent No. 6262233
; GENERAL INFORMATION:
; APPLICANT: GENTZ, REINER
; TITLE OF INVENTION: TISSUE FACTOR PATHWAY INHIBITOR-3
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, STE. 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,896A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-013-896A-14

Query Match 29.8%; Score 260; DB 4; Length 273;
Best Local Similarity 97.0%; Pred. No. 1e-57;
Matches 260; Conservative 0; Mismatches 8; Indels 0;
Gaps 0;
Qy 571 GACTCTCTCTCGCATGGCTGAGTGTGAGAGAGCCCGAGTGTCTCTGCTGATCG 630
Db 273 GACTCTCTCTCGCATGGCTGAGTGTGAGAGAGCCCGAGTGTCTCTGCTGATCG 214
Qy 631 ATTCTTCTCTCCAGGTAGAGTCTTCTGTTATGTTGAATTCATGCTCTTTTCTC 690
Db 213 ATTCTTCTCTCCAGGTAGAGTCTTCTGTTATGTTGAATTCATGCTCTTTTCTC 154
Qy 691 ATCAAGAGTATGATGTGAATCGTTCTTTGTTGCTGATTAAGTCTTTTAAGT 750

```

Db 153 ATCCAGAGATGATGTGGAAATCGTTCTTTGTTGCTGATTAATGTTTTTAAGT 94
Qy 751 ATAAACAAAGTTTATTTAGCATTCCTGAAGAGAAAGTAATGCAAGTTTAAT 810
Db 93 ATAAACAAAGTTTATTTATNNATTCCTGAAGAGAAAGTAATGCAAGTTTAAT 34
Qy 811 AAAAAGGGGCTTCCCTTTAGAAATAAA 838
Db 33 ANAAGGGGCTTCCCTTTAGAAATAAA 6

RESULT 7
US-09-013-896A-15/c
; Sequence 15, Application US/09013896A
; Patent No. 6262233
; GENERAL INFORMATION:
; APPLICANT: GENTZ, REINER
; TITLE OF INVENTION: TISSUE FACTOR PATHWAY INHIBITOR-3
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, STE. 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,896A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1290001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 256 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-013-896A-15

Query Match 28.6%; Score 250; DB 4; Length 256;
Best Local Similarity 99.6%; Pred. No. 3,6e-55;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 588 CCTGACAGTCGAGAGAGCCCGAGTGTCTCTGCTGATGATTTCTTCTCCAGGT 647
Db 256 CTTGACGTCGAGAGAGCCCGAGTGTCTCTGCTGATGATTTCTTCTCCAGGT 197
Qy 648 AGATTTTCTTTGCTATGTAATTCATTGCTTTTCTCATCAGAGAGATGTT 707
Db 196 AGATTTTCTTTGCTATGTAATTCATTGCTTTTCTCATCAGAGAGATGTT 137
Qy 708 GGAATGCTTTCTTTGTTGTCTGATTAATGTTTTTTAAGTATAACAAAGTTTTT 767
Db 136 GGAATGCTTTCTTTGTTGTCTGATTAATGTTTTTTAAGTATAACAAAGTTTTT 77
Qy 768 ATTAGCATTCGAAGAGAAAGTAATGTAAGTTTATTAAGAGGGGCTTCCCTCC 827
Db 76 ATTAGCATTCGAAGAGAAAGTAATGTAAGTTTATTAAGAGGGGCTTCCCTCC 17
Qy 828 TTTAGATATAA 838

```

```

Db 16 TTTAGATATAA 6

RESULT 8
US-09-013-896A-17
; Sequence 17, Application US/09013896A
; Patent No. 6262233
; GENERAL INFORMATION:
; APPLICANT: GENTZ, REINER
; TITLE OF INVENTION: TISSUE FACTOR PATHWAY INHIBITOR-3
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, STE. 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,896A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1290001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-013-896A-17

Query Match 21.5%; Score 188; DB 4; Length 201;
Best Local Similarity 99.0%; Pred. No. 2,5e-39;
Matches 199; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 310 GAGGAACGAGAGCGTGCCTGCGACCGCTGAGAGCTCCGAGATGACAAGAGGAGCT 369
Db 1 GAGGAACGAGAGCGTGCCTGCGACCGCTGAGAGCTCCGAGATGACAAGAGGAGCT 60
Qy 370 GGTGAAGAACACATATGTCTGTGACCGCCCTGTGCGCAAGAGACT-GGGAAGGAGGG 428
Db 61 GGTGAAGAACACATATGTCTGTGACCGCCCTGTGCGCAAGAGACTGGGAAGGAGGG 120
Qy 429 GAGACTATGTGACCTTTTATTAATGAGGATTAAGTGGATTTGATGATCTTGG 488
Db 121 GAGACTATGTGACCTTTTATTAATGAGGATTAAGTGGATTTGATGATCTTGG 180
Qy 489 GGCTGAGTCTGTTCTCTGG 509
Db 181 GGCTGAGTCTGTTCTCTGG 201

RESULT 9
US-09-404-879A-169
; Sequence 169, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.

```



```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 169
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-404-879A-169

Query Match
19.2%; Score 167.4; DB 4; Length 276;
Best Local Similarity 99.4%; Pred. No. 5.6e-34;
Matches 168; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 60
DB 88 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 147
QY 61 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 120
DB 148 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 207
QY 121 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 169
DB 208 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 256

RESULT 10
US-09-404-879A-356
; Sequence 356, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 356
; LENGTH: 207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-404-879A-356

Query Match
19.1%; Score 167; DB 4; Length 207;
Best Local Similarity 100.0%; Pred. No. 6.2e-34;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 60
DB 41 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 100
QY 61 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 120
DB 101 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 160
QY 121 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 167
DB 161 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 207

RESULT 11
US-09-404-879A-365/c
; Sequence 365, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:

```

```

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 365
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-404-879A-365

Query Match
19.1%; Score 167; DB 4; Length 371;
Best Local Similarity 100.0%; Pred. No. 8.1e-34;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 60
DB 167 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 108
QY 61 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 120
DB 107 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 48
QY 121 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 167
DB 47 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 1

RESULT 12
US-09-404-879A-168/c
; Sequence 168, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 168
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(276)
; OTHER INFORMATION: n = A,T,C or G
US-09-404-879A-168

Query Match
18.9%; Score 165.4; DB 4; Length 276;
Best Local Similarity 98.2%; Pred. No. 1.8e-33;
Matches 166; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 60
DB 189 CTCACGATATGTTCAATATGAAAGATCTGACCCGCCAACGACGTCACTGGGCTTG 130
QY 61 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 120
DB 129 CCGTGATCCTTCCACGCTGTGACTTTGACGTGAGAGAACTCTGCAATACTTCAT 70
QY 121 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 169
DB 69 CTATGAGGCTGCCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCC 21

```


SEQUENCE CHARACTERISTICS:
; LENGTH: 783 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-030-607-15

Query Match 8.3%; Score 72.8; DB 4; Length 783;
Best Local Similarity 61.7%; Pred. No. 1.6e-09;
Matches 113; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 GAAGAACTACTGACCGCCCAACGACGATCTGGGCTTGGCGTGCATCTCTTCCACGCTGG 82
DB 252 GAAGACTACTGCTGCTGCTATCCACACANGTGGGTGCTGCGCGGCTCTTCCACGCTGG 311
QY 83 TACTTTGACGTGAGAGAGAACTCTTGCATTAATCTATGAGAGCTGCCGGGCAAT 142
DB 312 TACTATGACCCCAAGAGAGATCTGCACAGATTTGCTTATGAGAGCTGCTTGGCAAC 371
QY 143 AAGAACAGCTACCGCTCTGAGAGAGCTGCATGCTCCGCTTCCGACAGAGAAAT 202
DB 372 AAGAACACTACTCTTGGGAGAGAGATGATTTCTANCTGTONGGAGTGTGCACAGTGGG 431
QY 203 CCT 205
DB 432 CCT 434

Search completed: January 10, 2003, 03:22:17
Job time : 56.01 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 02:45:21 (Search time 199.668 Seconds)

(without alignments)
1924.247 Million cell updates/sec

Title: US-09-825-682a-56

Perfect score: 873

Sequence: CTCACGCGATGCTTCACT.....TCAAAAAAAAAAAAAA 873

Scoring table:

IDENTITY_NUC

Gapop 10.0, Gapept 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCNT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCNTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862	98.7	1610	US-09-827-948-1	Sequence 1, Appl
2	843	96.6	1527	US-09-822-830A-95	Sequence 95, Appl
3	825.2	94.5	1544	US-09-880-107-3429	Sequence 3429, Ap
4	758.2	86.8	1649	US-09-925-301-424	Sequence 424, App
5	316.4	36.2	320	US-09-604-287A-293	Sequence 293, App
6	316.4	36.2	320	US-09-339-338-293	Sequence 293, App
7	316.4	36.2	320	US-10-007-805-293	Sequence 293, App
8	312.2	35.8	331	US-09-869-708-166	Sequence 166, App
9	312.2	35.8	331	US-09-880-107-2048	Sequence 2048, Ap
10	288.4	33.0	320	US-09-867-701-510	Sequence 510, App
11	282.4	32.3	287	US-09-827-948-13	Sequence 13, Appl
12	273	31.3	285	US-09-815-343-32	Sequence 32, Appl
13	262	30.0	286	US-09-815-343-703	Sequence 703, Appl
14	260	29.8	273	US-09-827-948-14	Sequence 14, Appl
15	253.8	29.1	287	US-09-815-343-123	Sequence 15, Appl
16	250	28.6	256	US-09-827-948-15	Sequence 15, Appl
17	238.4	27.3	241	US-09-604-287A-376	Sequence 376, App
18	238.4	27.3	241	US-10-007-805-376	Sequence 376, App
19	226.4	25.9	267	US-09-815-343-1258	Sequence 1258, Ap

20	188	21.5	201	US-09-827-948-17	Sequence 17, Appl
21	186.2	21.3	396	US-09-970-966-18	Sequence 18, Appl
22	186.2	21.3	396	US-09-825-294-18	Sequence 18, Appl
23	167.4	19.2	276	US-09-884-441-169	Sequence 169, App
24	167	19.1	207	US-09-884-441-356	Sequence 356, App
25	167	19.1	224	US-09-815-343-1343	Sequence 1343, Ap
26	167	19.1	234	US-09-815-343-1063	Sequence 1063, Ap
27	167	19.1	234	US-09-815-343-1129	Sequence 1129, Ap
28	167	19.1	234	US-09-815-343-1191	Sequence 1191, Ap
29	167	19.1	371	US-09-884-441-355	Sequence 355, App
30	165.4	18.9	234	US-09-815-343-1190	Sequence 1190, Ap
31	165.4	18.9	276	US-09-884-441-168	Sequence 168, App
32	150.6	17.3	302	US-09-815-343-991	Sequence 991, App
33	146.8	16.8	199	US-09-604-287A-125	Sequence 125, App
34	146.8	16.8	199	US-09-339-338-125	Sequence 125, App
35	146.8	16.8	199	US-10-007-805-125	Sequence 125, App
36	99.8	11.4	396	US-09-970-966-75	Sequence 75, Appl
37	99.8	11.4	396	US-09-825-294-75	Sequence 75, Appl
38	96.6	11.1	425	US-09-960-352-6197	Sequence 6197, Ap
39	95.4	10.9	369	US-09-960-352-12098	Sequence 12098, A
40	93.8	10.7	425	US-09-960-352-4987	Sequence 4987, Ap
41	72.8	8.3	783	US-09-232-880-15	Sequence 15, Appl
42	72.8	8.3	783	US-10-012-896-15	Sequence 15, Appl
43	72.8	8.3	783	US-09-895-793-15	Sequence 15, Appl
44	72.8	8.3	783	US-09-895-814-15	Sequence 15, Appl
45	72.8	8.3	783	US-09-759-143-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-827-948-1
Sequence 1, Application US/09827948
Patent No. US20010029034A1
GENERAL INFORMATION:
APPLICANT: Hsu, Tsu-An
APPLICANT: Rosen, Craig A.
APPLICANT: Ni, Jian
TITLE OF INVENTION: Tissue Factor Pathway Inhibitor-3
FILE REFERENCE: 1488.1290002
CURRENT APPLICATION NUMBER: US/09/827, 948
CURRENT FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: US 09/013, 896
PRIOR FILING DATE: 1998-01-27
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 1610
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (361)..(1116)
NAME/KEY: sig_peptide
LOCATION: (361)..(439)
NAME/KEY: mat_peptide
LOCATION: (442)..(1116)
US-09-827-948-1

Query Match 98.7%; Score 862; DB 10; Length 1610;
Best Local Similarity 99.9%; Pred. No. 1.6e-203;
Matches 873; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CTCACGCGATGCTTCACTGAAAGAACTAGACCGCAACGACGACGCGGCTTG 60
DB CTCACGCGATGCTTCACTGAAAGAACTAGACCGCAACGACGACGCGGCTTG 785
QY 61 CCGTCATCCTCCACGCGTGTGACTTGAAGTGAAGAGAACTCTGCAATTAATTCTAT 120
DB CCGTCATCCTCCACGCGTGTGACTTGAAGTGAAGAGAACTCTGCAATTAATTCTAT 845

QY	121	CTATGAGAGCTGCCTGGGSCATATAAGAACAGCTACCGCTGAGAGAGGCTCAGTCG	180
Db	846	CTATGAGAGCTGCCGGGSCATATAAGAACAGCTACCGCTGAGAGAGGCTCAGTCG	905
QY	181	CTGCTTCGGCAGAGAAATCCCTCCCTGCGCTCAAGAGTGAGTTCGCG	240
Db	906	CTGCTTCGGCAGAGAAATCCCTCCCTGCGCTCAAGAGTGAGTTCGCG	965
QY	241	GAGGCTGTGTGTGATGTGTGATCTCTTCTGTGGAGCTCCATGTGTACTGATCCG	300
Db	966	GAGGCTGTGTGTGATGTGTGATCTCTTCTGTGGAGCTCCATGTGTACTGATCCG	1025
QY	301	GGTGGCAGAGGAATCAGAGAGGTGCGCTGTGGCAACGTGTGGAGCTCCGAGATACAA	360
Db	1026	GGTGGCAGAGGAATCAGAGAGGTGCGCTGTGGCAACGTGTGGAGCTCCGAGATACAA	1085
QY	361	GGAGCAGCTGTGTAAACAACAATATGTCTGTGACCCGCTGTGGCAAGAGACT	419
Db	1086	GGAGCAGCTGTGTAAACAACAATATGTCTGTGACCCGCTGTGGCAAGAGACT	1145
QY	420	AAGSAGGGAGACTATGTGTGAGCTTTTTTAAATAGAGSAGTTGACTCGGATTTGAGT	479
Db	1146	AAGSAGGGAGACTATGTGTGAGCTTTTTTAAATAGAGSAGTTGACTCGGATTTGAGT	1205
QY	480	GATCATTAAGGCTGAGAGCTCTTCTCTGGAGAGTAGAGAGGCTTCTGTGTGGCA	539
Db	1206	GATCATTAAGGCTGAGAGCTCTTCTCTGGAGAGTAGAGAGGCTTCTGTGTGGCA	1265
QY	540	GGGATGGGTTTGTCTTGGAAATCCTCAGAGAGCTCCCTCGGCAATGGCTGAGCTGG	599
Db	1266	GGGATGGGTTTGTCTTGGAAATCCTCAGAGAGCTCCCTCGGCAATGGCTGAGCTGG	1325
QY	600	CAGCAGCCCGAGTTGTTTCTCGCTGATGCAATTTCTTCTCGAGTAGAGTTTCTT	659
Db	1326	CAGCAGCCCGAGTTGTTTCTCGCTGATGCAATTTCTTCTCGAGTAGAGTTTCTT	1385
QY	660	GCTTATGTGAATTCATTCGCTCTTTTCTCATCAAGAGTAGTGTGAATGCTTCT	719
Db	1386	GCTTATGTGAATTCATTCGCTCTTTTCTCATCAAGAGTAGTGTGAATGCTTCT	1445
QY	720	TTTGTGTGTGATTAAGGTTTTTAAAGTAAACAAGTTTTTATATAGATTCG	779
Db	1446	TTTGTGTGTGATTAAGGTTTTTAAAGTAAACAAGTTTTTATATAGATTCG	1505
QY	780	AAGAAGAGAAAGTAAATGTACAAGTTTAAATAAAGGGGCTTCCCTCTAGATTAAT	839
Db	1506	AAGAAGAGAAAGTAAATGTACAAGTTTAAATAAAGGGGCTTCCCTCTAGATTAAT	1565
QY	840	TTTACGATGTGCTTTCAAAAAAAAAAAAAAAAAAAAAA	873
Db	1566	TTTACGATGTGCTTTCAAAAAAAAAAAAAAAAAAAAAA	1599
RESULT 2			
US-09-822-830A-95			
Sequence 95, Application US/09822830A			
Patent No. US20020142952A1			
GENERAL INFORMATION:			
APPLICANT: Genetics Institute, Inc.			
APPLICANT: Mong, Gordon G.			
APPLICANT: Clark, Hilary			
APPLICANT: Fechtel, Kim			
APPLICANT: Agostino, Michael J.			
APPLICANT: Howes, Steven H.			
APPLICANT: Resnick, Richard J.			
APPLICANT: Gulukota, Kamalakat			
APPLICANT: Graham, James R.			
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS			
FILE REFERENCE: GIN 6402			
CURRENT APPLICATION NUMBER: US/09/822, 830A			
CURRENT FILING DATE: 2001-03-29			
PRIORITY APPLICATION NUMBER: 60/195, 604			
PRIORITY FILING DATE: 2000-04-06			

```

:      NUMBER OF SEQ ID NOS: 631
:      SOFTWARE: PatentIn Ver. 2.0
:      SEQ ID NO 95
:      LENGTH: 1527
:      TYPE: DNA
:      ORGANISM: Homo sapiens
:      US-09-822-830A-95

Query Match      96.6%; Score 843; DB 10; Length 1527;
Best Local Similarity 99.9%; Pred. No. 7.7e-199;
Matches 854; Conservative 0; Mismatches 0; Indels 1; Gaps 1

```

Query Match	96.6%;	Score 843;	DB 10;	Length 1527;
Best Local Similarity	99.9%;	Prod. No. 7.7e-199;		
Matches 854;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps
Qy 1 CTCACGATATGTTCACTATGAGAAATCTGCAACCGCAACGCAAGTCACTGAGCCTTG 60				
Dp 673 CTCACGAGATATGTTCACTATGAGAAATCTGCAACCGCAACGCAAGTCACTGAGCCTTG 732				
Qy 61 CCGTGCATCTTCCACCCCTGTGTAATCTTGACGTGAGAGAACTCTTGAAATTAATTAT 120				
Dp 733 CCGTGCATCTTCCACCCCTGTGTAATCTTGACGTGAGAGAACTCTTGAAATTAATTAT 792				
Qy 121 CTATGAGAGCTGCGGAGCAATAAGAAACGCTACCGCTGAGAGAGAGCTCATAGTCTCG 180				
Dp 793 CTATGAGAGCTGCGGAGCAATAAGAAACGCTACCGCTGAGAGAGAGCTCATAGTCTCG 852				
Qy 181 CTGCTTCGCAAGCAAGAAATCTCCCTGCCCCCTTGCTGCTGCTGAAAGTGTGTTGTGCG 240				
Dp 853 CTGCTTCGCAAGCAAGAAATCTCCCTGCCCCCTTGCTGCTGCTGAAAGTGTGTTGTGCG 912				
Qy 241 GGGGCTGTTGTGATAGTGTGTTGATCTCTTCTCTGAGAGCTCATAGTCTTAACCTGATCCG 300				
Dp 913 GGGGCTGTTGTGATAGTGTGTTGATCTCTTCTCTGAGAGCTCATAGTCTTAACCTGATCCG 972				
Qy 301 GGTGCGACAGAGGAACACAGAGAGGTGCGCTCGACCGCTGTGAGACTCCGAGATATACA 360				
Dp 973 GGTGCGACAGAGGAACACAGAGAGGTGCGCTCGACCGCTGTGAGACTCCGAGATATACA 1032				
Qy 361 GGAGCAGCTGTGTAAACAACATATATCTCTGTGACCGCTGTGCCCAAGAGACT -GAG 419				
Dp 1033 GGAGCAGCTGTGTAAACAACATATATCTCTGTGACCGCTGTGCCCAAGAGACT -GAG 1092				
Qy 420 AAGGAGGGGAGACATATGTGTGACCTTTTAAATGAGGGATTGACTCGATTTGACT 479				
Dp 1093 AAGGAGGGGAGACATATGTGTGACCTTTTAAATGAGGGATTGACTCGATTTGACT 1152				
Qy 480 GATCATTAAGGCTGAGAGTGTGTTCTCTGAGAGAGTGAAGACGGTGTCCCTCGGTCTGCGA 539				
Dp 1153 GATCATTAAGGCTGAGAGTGTGTTCTCTGAGAGAGTGAAGACGGTGTCCCTCGGTCTGCGA 1212				
Qy 540 GGGATGAGTGTGCTGTGGAATTCCTCTAGAGAGCTCTCTCTGCAATGAGCTGAGCTGAG 599				
Dp 1213 GGGATGAGTGTGCTGTGGAATTCCTCTAGAGAGCTCTCTCTGCAATGAGCTGAGCTGAG 1272				
Qy 600 CAGCAGCCCCGAGCTGTTCTCTGCTGATGATGATTTCTTCTCTCAGGTAGATTTTCTT 659				
Dp 1273 CAGCAGCCCCGAGCTGTTCTCTGCTGATGATGATTTCTTCTCTCAGGTAGATTTTCTT 1332				
Qy 660 GGTATGTTGAATTCATATGCTCTCTTCTCTGATCAACAGAGAGATGTGGATGCTTCT 719				
Dp 1333 GGTATGTTGAATTCATATGCTCTCTTCTCTGATCAACAGAGAGATGTGGATGCTTCT 1392				
Qy 720 TTGTGTGCTGATTTATGATTTATTTTAAAGTAAACAAAGTTTTTATATAGATCTG 779				
Dp 1393 TTGTGTGCTGATTTATGATTTTATTTAAGTATTAACAAAGTTTTTATATAGATCTG 1452				
Qy 780 AAAGAGAGAAATGAAATGTACAAAGTTATATATAAAGGGGCTTCCCTTTAGATTAAT 839				
Dp 1453 AAAGAGAGAAATGAAATGTACAAAGTTATATATAAAGGGGCTTCCCTTTAGATTAAT 1512				
Qy 840 TTCAGCATGTGCTTT 854				
Dp 1513 TTCAGCATGTGCTTT 1527				

RESULT 3

US-09-880-107-3429
 ; Sequence 3429, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darcil T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scheff, Uwe
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 3429
 ; LENGTH: 1544
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U78035
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(1535)
 ; OTHER INFORMATION: n = a or c or g or t
 US-09-880-107-3429

Query Match 94.5%; Score 825.2; DB 10; Length 1544;
 Best Local Similarity 98.5%; Pred. No. 1,9e-194;
 Matches 864; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

```

QY 1 CTCAGGATATGTTCACTATGAAGAATCTGCAACGCGCAACGAGTCACTGAGCCTTG 60
DB 666 CTCAGGATATGTTCACTATGAAGAATCTGCAACGCGCAACGAGTCACTGAGCCTTG 725
QY 61 CCGTGATCTTCCACGCTGCTGTAATTGACGTGAGAGAACTCTGCAATAAATTCTTCA 120
DB 726 CCGTGATCTTCCACGCTGCTGTAATTGACGTGAGAGAACTCTGCAATAAATTCTTCA 785
QY 121 CTATGAGAGCTCCGGGGCAATAGAACAGTACCGCTCTGAGAGAGGCTGCAATGCTCCG 180
DB 786 CTATGAGAGCTCCGGGGCAATAGAACAGTACCGCTCTGAGAGAGGCTGCAATGCTCCG 845
QY 181 CTGCTCCGCGACAGAGAAATCCCTCCGCTGCTGCTCAAGGTGAGTCTGAGC 240
DB 846 CTGCTCCGCGACAGAGAAATCCCTCCGCTGCTGCTCAAGGTGAGTCTGAGC 905
QY 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 906 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 965
QY 301 GGTGGCAGGAGAACCAAGAGCGTGCCTGCGACCGTCTGAGACTCCGAGATGACAA 360
DB 966 GGTGGCAGGAGAACCAAGAGCGTGCCTGCGACCGTCTGAGACTCCGAGATGACAA 1025
QY 361 GGAGCAGCTGTGAAGAACATATGTCGTGACCGCTGCGACCGTCTGAGACTCCGAGATGACAA 419
DB 1026 GGAGCAGCTGTGAAGAACATATGTCGTGACCGCTGCGACCGTCTGAGACTCCGAGATGACAA 1085
QY 420 AAGGAGAGGAGAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 479
DB 1086 AAGGAGAGGAGAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1145
QY 480 GATCAATTAAGGCTGAGGCTGTTTCTCTGAGAGGTGAGAGCGTCTGCTCTGCTGCTGCA 539
DB 1146 GATCAATTAAGGCTGAGGCTGTTTCTCTGAGAGGTGAGAGCGTCTGCTCTGCTGCTGCA 1205
QY 540 GGGATGGGTTGCTTTGAAATCCCTATGAGAGGCTCTCTCTGAGAGGCTGAGAGGCTGAG 599
DB 1206 GGGATGGGTTGCTTTGAAATCCCTATGAGAGGCTCTCTCTGAGAGGCTGAGAGGCTGAG 1265

```

```

QY 600 CAGCAGCCCCAGTGTGTTCTCTGATGATGATTTCTTCTCCAGGTAGATTTCTT 659
DB 1266 CAGCAGCCCCAGTGTGTTCTCTGATGATGATTTCTTCTCCAGGTAGATTTCTT 1325
QY 660 GCTTATGTTGAATTCATGCTC--TTTTCATCAGAGAGTGTGATGCTTT 717
DB 1326 GCTTATGTTGAATTCATGCTCCTTTCTCAGAGAGTGTGATGCTTT 1385
QY 718 CTTTGTGTTGCTGATTTATGTTTTTTTAAATTAACAAAGTTTATTAAGCATTC 777
DB 1386 CTTTGTGTTGCTGATTTATGTTTTTTTAAATTAACAAAGTTTATTAAGCATTC 1445
QY 778 TGAAGAGAGAAAGTAAATGTCACAGTTTAAATTAAGAGGCTCTCCCTTAGAATTA 837
DB 1446 TGAAGAGAGAAAGTAAATGTCACAGTTTAAATTAAGAGGCTCTCCCTTAGAATTA 1505
QY 838 ATTT-CAGCATGCTTCAAAAAAAAAAAAAA 873
DB 1506 ATTTCCAGCATGCTTCTTCAAAAAAAAAAAAAA 1542

```

RESULT 4

US-09-925-301-424
 ; Sequence 424, Application US/09925301
 ; Patent No. US2002052308A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA106
 ; CURRENT APPLICATION NUMBER: US/09/925,301
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05882
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1694
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 424
 ; LENGTH: 1649
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-301-424

Query Match 86.8%; Score 758.2; DB 10; Length 1649;
 Best Local Similarity 95.6%; Pred. No. 7.2e-178;
 Matches 833; Conservative 0; Mismatches 33; Indels 5; Gaps 5;

```

QY 1 CTCAGGATATGTTCACTATGAAGAATCTGCAACGCGCAACGAGTCACTGAGCCTTG 60
DB 784 CTCAGGATATGTTCACTATGAAGAATCTGCAACGCGCAACGAGTCACTGAGCCTTG 843
QY 61 CCGTGATCTTCCACGCTGCTGTAATTGACGTGAGAGAACTCTGCAATAAATTCTTCA 120
DB 844 CCGTGATCTTCCACGCTGCTGTAATTGACGTGAGAGAACTCTGCAATAAATTCTTCA 903
QY 121 CTATGAGAGCTCCGGGGCAATAGAACAGTACCGCTCTGAGAGGCTGCAATGCTCCG 180
DB 904 CTATGAGAGCTCCGGGGCAATAGAACAGTACCGCTCTGAGAGGCTGCAATGCTCCG 963
QY 181 CTGCTCCGCGACAGAGAAATCCCTCCGCTGCTGCTCAAGGTGAGTCTGAGC 240
DB 964 CTGCTCCGCGACAGAGAAATCCCTCCGCTGCTGCTCAAGGTGAGTCTGAGC 1023
QY 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 1024 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1083
QY 301 GGTGGCAGGAGAACCAAGAGCGTGCCTGCGACCGTCTGAGACTCCGAGATGACAA 360
DB 1084 GGTGGCAGGAGAACCAAGAGCGTGCCTGCGACCGTCTGAGACTCCGAGATGACAA 1143
QY 361 GGAGCAGCTGTGAAGAACATATGTCGTGACCGCTGCGACCGTCTGAGACTCCGAGATGACAA 420

```

Db	1144	GGAGCAGCTGTTAAACAACATATGTCCTGG-CCGCGCTCGCAAGAGACTGAGG	1202
Qy	421	AGGAGGGGAGACTATGCTGTAGCTTTTTTAAATAGAGGGATTGACCTCGATTGAGTG	480
Db	1203	AAAGGAGGGAGACATGTGTGA-CTTTTTTAAATAGAGGATGACTCGGATTTGAGTG	1261
Qy	481	ATCATTAAGGCTAGAGCTCTGTTCTCGGAGAGTAGAGAGCTGCTCTGATCTGAGCAG	540
Db	1262	ATCATTAAGGCTAGAGCTGTTCTCGGAGAGTAGAGAGCTGCTCTGATCTGAGCAG	1321
Qy	541	GGATGGGTTGCTTTTGAAATCCTCTAGAGGCTCTCTCGATAGGCTCGAGCTGCGC	600
Db	1322	GGATGGGTTGCTTTTGAAATCCTCTAGAGGCTCTCTCGATAGGCTCGAGCTGCGC	1381
Qy	601	AGCAGCCCCGAGTGGTTTCTCGATGATGATTTTCTTCGAGGTAGATTTCTTTG	660
Db	1382	AGCAGCCCCGAGTGGTTTCTCGATGATGATTTTCTTCGAGGTAGATTTCTTTG	1440
Qy	661	CTTATGTTGAATTCATATGCTCTTTCTCTCATCAGAGTGTGTTGGAATCGTTCTT	720
Db	1441	CTTATGTTGA-TCATATGCTCTTTCTCTCATCAGAGTGTGTTGGAATCGTTCTT	1499
Qy	721	TTGTTGTGCAATTATGTTTTTTAATGATTAACAAAGTTTTTATATGACATCTGA	780
Db	1500	TTGTTGTGCAATTATGTTTTTTAATGATTAACAAAGTTTTTATATGACATCTGA	1559
Qy	781	AGAGAGGAAGTAAATGTACAGATTAAATAAAGAGGAGCTTCCCTTTAGAAATAAT	840
Db	1560	AGAGAGGAAGT-AAATGTACAGATTAAATAAAGAGGAGCTTCCCTTTATATAAAAAA	1618
Qy	841	TCAGATGTGCTTCAAAAAAAAAAAAAA 871	
Db	1619	AAAAAAAAAAAAAAAAAAAAAAAAAAAA 1649	

	RESULT 5	
	US-09-604-287A-293	
	: Sequence 293, Application US/09604287A	
	: Patent No. US20020064872A1	
	: GENERAL INFORMATION:	
	: APPLICANT: Jiang, Yugu	
	: APPLICANT: Dillon, Davin C.	
	: APPLICANT: Mitcham, Jennifer L.	
	: APPLICANT: Xu, Jianguan	
	: APPLICANT: Harlocker, Susan L.	
	: APPLICANT: Hepler, William T.	
	: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND	
	: TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER	
	: FILE REFERENCE: 210121.47007	
	: CURRENT APPLICATION NUMBER: US/09/604,287A	
	: CURRENT FILING DATE: 2000-06-22	
	: NUMBER OF SEQ ID NOS: 489	
	: SOFTWARE: PastSeq for Windows Version 3.0	
	: SEQ ID NO 293	
	: LENGTH: 320	
	: TYPE: DNA	
	: ORGANISM: Homo sapiens	
	US-09-604-287A-293	
	Query Match	36.2%; Score 316.4; DB 10; Length 320;
	Best Local Similarity	99.7%; Pred. No. 4.ee-69;
	Matches 317; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Oy	519 CGCGCTGCTCCGGGTGGAGGATGGGATTGCTTGGAAACCTCTGAGAGGCTCTC	578
Dd	2 CGGCTCTTCCTGGTCTGGGCGGGAAGGGTTTGCTTTGGAAATCTCTTGGAGGCTCTC	61
Oy	579 CTGCAGTAGGCGCTGCACTCTGGACAGACCCTCGAGTTGTTCTCTGCTGATCATTTCTTT	638
Dd	62 CTCGAGATGGCCCTGGAGCTCTGGACAGACGCCCGAGTGTCTTCCTGCTGATCATTTCTTT	121
Oy	639 CTCACAGTAGAGTTTCTTCTTACTTATGTGTAATTCATTCCTCTTTTCTCATACAGA	698

Db	122	CCGCCAGTAGAGTTTCTTTCGTTATGTTGATTCGATTCGATTCGCTCTTCTCTCATCAGCA	181
Qy	699	AGTGATGTTGGAATCGTTCTTTTGTGTTGTCTGANTTAAGTTTTTTAAGATPAAACA	758
Db	182	AGGATGTTGGAAATCGTTCTTTCTTTGTTGTCGATTTAAGTTTTTTAAGTATPAAACA	244
Qy	759	AGGTTTTTATTTAGCATCTCGAAGAGAGAAAGAACTAAATATGTAAGTTATATPAAAAGG	818
Db	242	AGGTTTTTATTTAGCATCTCGAAGAGAGAAAGAACTAAATATGTAAGTTATATPAAAAGG	301
Qy	819	GCCTTCCCTTTAGAATA	836
Db	302	GCCTTCCCTTTAGAATA	319

RESULT 6
 US-09-339-338-293
 : Sequence 293, Application US/0933338A
 : Patent No. US20020102602A1
 : GENERAL INFORMATION:
 : APPLICANT: Yuguo, Jiang
 : APPLICANT: Dillon, David C.
 : APPLICANT: Mitcham, Jennifer L.
 : APPLICANT: Xu, Jiangchun
 : TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
 : FILE REFERENCE: 210121.470C2
 : CURRENT APPLICATION NUMBER: US/09/339,338A
 : CURRENT FILING DATE: 1999-06-23
 : NUMBER OF SEQ. ID NOS: 315
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ. ID NO. 293
 : LENGTH: 320
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : US-09-339-338-293

```

Query Match Similarity      36.2%; Score 316.4; DE 10; Length 320;
Best Local Similarity      99.7%; Pred. No. 4.ee-69;
Matches 317; Conservative  0; Mismatches  1; Indels  0; Gaps  0

QY      519  CGCGTCGCTCTCTGAGTCGAGCAGGATGAGGTTGCTTTGAAATCTCTAGAGAGCTCTCTC 578
Db      2    CGCGTCGCTCTCTGCTCTCGCGGAGATGGGTTGCTTTGAAATCTCTAGAGAGCTCTCTC 61

QY      579  CTGCGATAGGCGCGAGAGTCGTGGACGACGCCCGAGTGTCTTCCTCGCTGATCATTTCTTT 638
Db      62  CTGCGATGCGCTGAGAGTCGTGGACGACGCCCGAGTGTCTTCCTCGCTGATCATTTCTTT 121

QY      639  CCTCCAGTAGAGATTTCCTTGGCTTATGTAAATTCATTCGCTCTTTCCTCATCAGAGA 698
Db      122  CCTCCAGTAGAGATTTCCTTGGCTTATGTAAATTCATTCGCTCTTTCCTCATCAGAGA 181

QY      699  AAGTAGATGTGAATACGCTTCTTGTGTGTCTGATTAATAGTTTTTAAAGTAAACAA 758
Db      182  AAGTAGATGTGAATACGCTTCTTGTGTGTCTGATTAATAGTTTTTAAAGTAAACAA 241

QY      759  AAGTTTTTTATTAGCATTCGAAGAAGAGAAAGTAAATGTACAACTTTAATTAATAAAAAGG 818
Db      242  AAGTTTTTTATTAGCATTCGAAGAAGAGAAAGTAAATGTACAACTTTAATTAATAAAAAGG 301

QY      819  GCCTTCCCTTTAGAATA 836
Db      302  GCCTTCCCTTTAGAATA 319

RESULT 7
US-10-007-805-293
; Sequence 293, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu

```



```
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margareta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 293
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-007-805-293
```

```
Query Match          36.2%; Score 316.4; DB 12; Length 320;
Best Local Similarity 99.7%; Pred. No. 4,6e-69;
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 519 CGGCTCTTCTGCTGCTGCGAGGAGTGGTTGCTTTGGAAATCCTCTAGAGAGCTTCCTC 578
DB 2 CGGCTCTTCTGCTGCTGCGAGGAGTGGTTGCTTTGGAAATCCTCTAGAGAGCTTCCTC 61
QY 579 CTCGATGAGCTGCTGCTGCGAGGAGTGGTTGCTTTGCTGCTGCTGCTGCTGCTGCTGCT 638
DB 62 CTCGATGAGCTGCTGCTGCGAGGAGTGGTTGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 121
QY 639 CTCGATGAGTGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
DB 122 CTCGATGAGTGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181
QY 699 AGTGAAGTGGAAATGCTTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 758
DB 182 AGTGAAGTGGAAATGCTTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 241
QY 759 AAGTTTATATAGATCTGAAGAAGAAAGAAAGTAAATGTAAGTTTAAAGTAAAGG 818
DB 242 AAGTTTATATAGATCTGAAGAAGAAAGTAAATGTAAGTTTAAAGTAAAGG 301
QY 819 GCCTTCCCTTTAGATA 836
DB 302 GCCTTCCCTTTAGATA 319
```

RESULT 8

```
US-09-969-708-166/c
; Sequence 166, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 166
; LENGTH: 331
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: n=a,t,g or c
US-09-969-708-166
```

```
Query Match          35.8%; Score 312.2; DB 10; Length 331;
Best Local Similarity 98.2%; Pred. No. 5,1e-68;
Matches 325; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
```

```
QY 484 ATTAGAGCTGAGTCTGTTTCTGAGAGTAGACGCTG-CTTCTGCTGCTGAGAGG 542
DB 331 ATTAGAGCTGAGTCTGTTTCTGAGAGTAGACGCTGCTTCTGCTGCTGAGAGG 272
QY 543 ATGGTTTGTCTTGAATCTCTAGAGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
DB 271 ATGGTTTGTCTTGAATCTCTAGAGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 212
QY 603 CAGCCCGAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 662
DB 211 CAGCCCGAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 152
QY 663 TATGTTGAATTCATTCATTCCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722
DB 151 TATGTTGAATTCATTCATTCCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 92
QY 723 GTTTGCTGATTTAAGTTTCTTTTAAATTAACAAAGTTTCTTTTATTAAGATTCGAA 782
DB 91 GTTTGCTGATTTAAGTTTCTTTTAAATTAACAAAGTTTCTTTTATTAAGATTCGAA 32
QY 783 GAAGGAAATTAAGTACAAAGTTTAAATAA 813
DB 31 GAAGGAAATTAAGTACAAAGTTTAAATAA 1
```

RESULT 9

```
US-09-880-107-2048/c
; Sequence 2048, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2048
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 H95233
; NAME/KEY: unsure
; LOCATION: (1)..(331)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-2048
```

```
Query Match          35.8%; Score 312.2; DB 10; Length 331;
Best Local Similarity 98.2%; Pred. No. 5,1e-68;
Matches 325; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
```

```
QY 484 ATTAGAGCTGAGTCTGTTTCTGAGAGTAGACGCTG-CTTCTGCTGCTGAGAGG 542
DB 331 ATTAGAGCTGAGTCTGTTTCTGAGAGTAGACGCTGCTTCTGCTGCTGAGAGG 272
```



```

: CURRENT FILING DATE: 2001-03-22
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: FastSeq for Windows Version 4.0.0
: SEQ ID NO 32
: LENGTH: 285
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-815-343-32

```

Query Match	31.3%	Score 273;	DB 10;	Length 285;
Best Local Similarity	99.6%;	Pred. No. 2.3e-58;		
Matches 284;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1

```

Db      181  TCGGAGATGACAAAGGAGCAGCTGGTGTGAAGAACAACATATGTCTCGTACACCGCCCTGTGCG 240
Qy      407  CAAAGAGACT-GGGAAAGGAGGGGAGACTATGTGTGAGCTTTTTT 451
          |||||
Db      241  CAAAGAGACTGGGGAAGGGAAGGAGACTATGTGTGAGCTTTTTT 286
          |||||

```

Qy	168	CCGCAATCCTCCGCGTTCGCGACGACGAGATCCTCCCTGCCCTTGGCTCAAAAG	22
Db	285	CCGCAATCCTCCGCGTTCGCGACGACGAGATCCTCCCTGCCCTTGGCTCAAAAG	22
Qy	228	TGATGATCTGAGCGAGGCTGTTCGTGATGATGATGATCTCTTCTCTGGAACCTTCATAG	28
Db	225	TGATGATCTGAGCGAGGCTGTTCGTGATGATGATGATCTCTTCTCTGGAACCTTCATAG	16
Qy	288	TCTACCTATTCGAGGTGGACCGAGAGATCCAGAGAGCTGCTGGACACGTCGTGAAGCT	34
Db	165	TCTACCTATTCGAGGTGGACCGAGAGATCCAGAGAGCTGCTGGACACGTCGTGAAGCT	10
Qy	348	CCGAGATGACAAAGACAGCTGGTGAAGAACATATGTCTGTGACCGGCTGTGGCC	40
Db	105	CCGAGATGACAAAGACAGCTGGTGAAGAACATATGTCTGTGACCGGCTGTGGCC	46
Qy	408	AAAGAGACT-GGAGAGGAGAGGAGACATATGTGAGACTTTT	451
Db	45	AAAGAGACTGGAGAGGAGAGGAGACATATGTGAGACTTTTT	1

```

RESULT 13
US-09-815-343-703
: Sequence 703, Application US/09815343
: Patent No. US20010055596A1
: GENERAL INFORMATION:
: APPLICANT: Weagher, Madeleine
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.504
: CURRENT APPLICATION NUMBER: US/09/815,343
: CURRENT FILING DATE: 2001-03-22
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 703
: LENGTH: 286
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-815-343-703

```

Query Match	Best Local Similarity	Score	262; DE 10;	Length	286;
Matches	284; Conservative	0; Mismatches	0; Indels	2; Gaps	
Qy	168	CCGCGATGCTCCGCTGCTTTCGCCGACGAGGAAATCTCCCTGCTCCCTTGAGCTCAAGG	22		
Db	1	CCGCGATGCTCCGCTGCTTTCGCCGACGAGGAAATCTCCCTGCTCCCTTGAGCTCAAGG	60		
Qy	228	TGATGATCTCGACGGGCTGTTCGTGATGATGATGATCTCTTCCTCGGAGCCTCCATAG	28		
Db	61	TGATGATCTCGACGGGCTGTTCGTGATGATGATGATCTCTTCCTCGGAGCCTCCATAG	121		
Qy	288	TCTACCTGATCC-GGATGG-CAGGAGAGAAACAGAGAGTGGCCCTGSCGACCGCTGAGAC	34		
Db	121	TCTACCTGATCCGGGATGGACGAGAACACAGAGAGTGGCCCTGSCGACCGCTGAGAC	181		
Qy	347	TCCGAGATGACAGGAGCAGCTGGTGAAGAACATATGTCTCTGTACGGCCCTGTGCC	40		

RESULT 14

```

; Sequence 14, Application US/09827948
; Patent No. US20010029034A1
; GENERAL INFORMATION:

```

```

1  APPLICANT:  Hsu, Tsu-an
2  APPLICANT:  Rosen, Craig A.
3  APPLICANT:  Ni, Jian
4  TITLE OF INVENTION:  Tissue Factor Pathway Inhibitor-3
5  FILE REFERENCE:  1488.1290002
6  CURRENT APPLICATION NUMBER:  US/09/827,948
7  CURRENT FILING DATE:  2001-04-06
8  PRIOR APPLICATION NUMBER:  US/09/023,896
9  PRIOR FILING DATE:  1998-01-27
10 NUMBER OF SEQ ID NOS:  31
11 SOFTWARE:  PatentIn version 3.0
12 SEQ ID NO 14
13 LENGTH:  273
14 TYPE:  DNA
15 ORGANISM:  Human
16 FEATURE:
17 NAME/KEY:  misc_feature
18 LOCATION:  (32)_feature
19 OTHER INFORMATION:  n is A, T, C, or G

```

	Query Match	29.8%	Score 260	DB 105	Length 273
	Best Local Similarity	97.0%	Pred. No. 3	7e-55	
	Matches 260	Conservative 0	Mismatches 8	Indels 0	Gaps
QY	571	GAGTCTCTCTGATGAGCCTGAGTCTGAGAGAGCCCCAGAGTTGTTCTCGCTGATGC	633		
Db	273	GAGTCTCTCTGAGATGAGCCTGAGAGCTGAGAGAGGCCCGAGTGTCTCTCGAGATGC	214		
QY	631	ATTCTTCTCTCCAGTAGAGTTTCTTGTATGTTGAATTCATTGCTCTTTTCTC	690		
Db	213	ATTCTTCTCTCCAGTAGAGTTTCTTGTATGTTGAATTCATTGCTCTTTTCTC	154		
QY	691	ATACAGAGAGGATGTGGATGCTTTCTTTGTGTCGATTTATGATGTTTTTAAGT	753		
Db	153	ATCAGAGAGGATGTGGATGCTTTCTTTGTGTCGATTTATGATGTTTTTAAGT	94		
QY	751	ATTAACAAAGTTTTTATTAGCATCTGAGAAAGAGAAAGTAAATATACAGTTTAT	810		

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:46:01 : Search time 1874.25 Seconds
(without alignments)
7543.621 Million cell updates/sec

Title: US-09-825-682a-56
Perfect score: 873
Sequence: 1 ctccagcgatagtcgaact.....tcaaaaaaaaaaaaaa 873

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esthum:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
C 1	803.2	92.0	917	9	AL576113	AL576113
C 2	802.8	92.0	910	9	AL572703	AL572703
C 3	785	89.9	1001	9	AL548687	AL548687
C 4	783.4	89.7	923	10	BE621676	BE621676
C 5	783.4	89.6	1061	9	AL575665	AL575665
C 6	781.6	89.5	983	9	AL573986	AL573986

C 7	776.2	88.9	938	9	AL575631	AL575631
C 8	776	88.9	1052	9	AL573346	AL573346
C 9	770.8	88.3	877	9	AL576440	AL576440
C 10	770.6	88.3	950	9	AL576831	AL576831
C 11	752.8	86.2	1062	9	AL575764	AL575764
C 12	748	85.7	964	9	AL560664	AL560664
C 13	747.8	85.7	859	9	AL582333	AL582333
C 14	742.2	85.0	990	9	AL571794	AL571794
C 15	738	84.5	750	14	BM977432	BM977432
C 16	735.6	84.3	985	9	AL551363	AL551363
C 17	735.2	84.2	889	14	BQ439520	BQ439520
C 18	734.6	84.1	810	9	AL186641	AL186641
C 19	734.6	84.1	942	12	BE741988	BE741988
C 20	734	84.1	951	9	AL536057	AL536057
C 21	730.2	83.6	814	9	AL1983859	AL1983859
C 22	716.6	82.1	1038	9	AL571204	AL571204
C 23	700.6	80.3	986	12	BG681770	BG681770
C 24	700.2	80.2	1143	14	BM924750	BM924750
C 25	692.2	79.3	938	14	BQ949537	BQ949537
C 26	689.4	79.0	726	14	BM979926	BM979926
C 27	689	78.9	713	14	BQ181625	BQ181625
C 28	685.4	78.5	728	14	BQ575516	BQ575516
C 29	681	78.0	746	12	BE743960	BE743960
C 30	677	77.5	789	13	BM044691	BM044691
C 31	675	77.3	958	13	B1523756	B1523756
C 32	673	77.1	758	12	BE744503	BE744503
C 33	672	77.0	745	9	AL1819257	AL1819257
C 34	668.4	76.6	949	12	BE741701	BE741701
C 35	667.4	76.4	880	14	BQ227721	BQ227721
C 36	665.8	76.3	716	9	AL574271	AL574271
C 37	663.2	76.0	719	10	BE613272	BE613272
C 38	663	75.9	908	12	BE747343	BE747343
C 39	662.4	75.9	748	10	AM190372	AM190372
C 40	659.4	75.5	703	9	AL1982697	AL1982697
C 41	654.8	75.0	951	14	BQ947368	BQ947368
C 42	649.6	74.4	727	13	B1087012	B1087012
C 43	649.2	74.4	802	9	AL572724	AL572724
C 44	645.4	73.9	840	12	BE740873	BE740873
C 45	644	73.8	668	14	BM978012	BM978012

ALIGNMENTS

RESULT 1
AL576113/c 917 bp mRNA linear EST 16-FEB-2001
LOCUS AL576113 LRT NFL006.PL2 Homo sapiens cDNA clone CSOD1072YC24 3
DEFINITION prime, mRNA sequence.
ACCESSION AL576113
VERSION AL576113.1 GI:12937934
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 917)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 917
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSOD1072YC24"
/clone_1ib="LRT NFL006.PL2"
/issue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI, 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>

BASE COUNT 243 a 254 c 216 g 200 t 4 others
ORIGIN

Query Match 92.0%; Score 803.2; DB 9; Length 917;
Best Local Similarity 99.4%; Pred. No. 1.4e-136;
Matches 813; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

```

QY 1 CTCACGCGATATGTTCAACTATGAAGAATCTGACCGCCAAACGCACTCATGCGGCTTG 60
DB 818 CTCACGCGATATGTTCAACTATGAAGAATCTGACCGCCAAACGCACTCATGCGGCTTG 759
QY 61 CCGTGCACTCTTCCACCGCTGTACTTTGACGTGAGAGAGAACTCTTGCATTAATTTCT 120
DB 758 CCGTGCACTCTTCCACCGCTGTACTTTGACGTGAGAGAGAACTCTTGCATTAATTTCT 699
QY 121 CTATGAGAGCTGCCGGGGCAATAAGAACACTACCGCTCTGAGAGAGCTCTGATCTCCG 180
DB 698 CTATGAGAGCTGCCGGGGCAATAAGAACACTACCGCTCTGAGAGAGCTCTGATCTCCG 639
QY 181 CTGCTTCCGCAAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGTGTCTTCCG 240
DB 638 CTGCTTCCGCAAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGTGTCTTCCG 579
QY 241 GGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 578 GGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 519
QY 301 GGTGGCAGGAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 518 GGTGGCAGGAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
QY 361 GAGACAGCTGTGTGAAGAACATATGTCCTGTGACCGCTCTGCAAGAGAGACT- GGG 419
DB 458 GAGACAGCTGTGTGAAGAACATATGTCCTGTGACCGCTCTGCAAGAGAGACT- GGG 399
QY 420 AAGGAGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGAGATTTGACTCGAATTGAGT 479
DB 398 AAGGAGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGAGATTTGACTCGAATTGAGT 339
QY 480 GATCATTTAGAGCTGAGAGCTCTTTCTCTGGAGAGTGAAGAGCGCTCTTCCGTCTGGCA 539
DB 338 GATCATTTAGAGCTGAGAGCTCTTTCTCTGGAGAGTGAAGAGCGCTCTTCCGTCTGGCA 279
QY 540 GGAATGGGTTTCTTGGAAATCCTCTAGAGAGCTCTCTCGTCAATGAGCTGAGTGG 599
DB 278 GGAATGGGTTTCTTGGAAATCCTCTAGAGAGCTCTCTCGTCAATGAGCTGAGTGG 219
QY 600 CAGCAGCCCGAGTTGTTCTCTGCTAGTCGATTTCTTCTCCAGAGTGAAGTTTCTTT 639
DB 218 CAGCAGCCCGAGTTGTTCTCTGCTAGTCGATTTCTTCTCCAGAGTGAAGTTTCTTT 159
QY 660 GCTTATGTTGAATTCATTTGCTCTTTCTATACAGAGAGTGTGGATGCTTTCT 719
DB 158 GCTTATGTTGAATTCATTTGCTCTTTCTATACAGAGAGTGTGGATGCTTTCT 99
QY 720 TTGTTGTCTGATTTATGTTTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 779
DB 98 TTGTTGTCTGATTTATGTTTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 39
QY 780 AAAGAGAGAAAGTAAATGTACAGTTTAAATTAAG 817
DB 38 AAAGAGAGAAAGTAAATGTACAGTTTAAATTAAG 1

```

RESULT 2

AL572703/c
LOCUS AL572703 910 bp mRNA linear EST 16-FEB-2001
DEFINITION AL572703 LTI NPL006 Pl2 Homo sapiens cDNA clone CS0D1035YE24 3
prime, mRNA sequence.
ACCESSION AL572703
VERSION AL572703.1 GI:12911228
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1035YE24"
/clone_id="LTI NPL006 Pl2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>"

BASE COUNT 244 a 251 c 214 g 195 t 6 others
ORIGIN

Query Match 92.0%; Score 802.8; DB 9; Length 910;
Best Local Similarity 99.3%; Pred. No. 1.4e-136;
Matches 812; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

```

QY 1 CTCACGATATGTTCAACTATGAAGAATCTGACCGCCAAACGCACTCATGCGGCTTG 60
DB 824 CTCACGATATGTTCAACTATGAAGAATCTGACCGCCAAACGCACTCATGCGGCTTG 765
QY 61 CCGTGCACTCTTCCACCGCTGTACTTTGACGTGAGAGAGAACTCTTGCATTAATTTCT 120
DB 764 CCGTGCACTCTTCCACCGCTGTACTTTGACGTGAGAGAGAACTCTTGCATTAATTTCT 705
QY 121 CTATGAGAGCTGCCGGGGCAATAAGAACACTACCGCTCTGAGAGAGCTCTGATCTCCG 180
DB 704 CTATGAGAGCTGCCGGGGCAATAAGAACACTACCGCTCTGAGAGAGCTCTGATCTCCG 645
QY 181 CTGCTTCCGCAAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGTGTCTTCCG 240
DB 644 CTGCTTCCGCAAGAGAGAAATCCCTCCCTGCTTGGCTCAAGGTGTGTCTTCCG 585
QY 241 GGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 584 GGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 525
QY 301 GGTGGCAGGAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 524 GGTGGCAGGAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465
QY 361 GAGACAGCTGTGTGAAGAACATATGTCCTGTGACCGCTCTGCAAGAGAGACT- GGG 419
DB 464 GAGACAGCTGTGTGAAGAACATATGTCCTGTGACCGCTCTGCAAGAGAGACT- GGG 405
QY 420 AAGGAGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGAGATTTGACTCGAATTGAGT 479

```

Db 404 AAGGAGGAGGAGACTATGTGTAGCTTTTAAATAGAGGATTGACTCGGATTTGACT 345
Qy 480 GATCATTAGGAGCTGAGGCTGTCTTCTCGGAGAGTAGAGACGCTGCTTCTGCTGGCA 539
Db 344 GATCATTAGGAGCTGAGGCTGTCTTCTCGGAGAGTAGAGACGCTGCTTCTGCTGGCA 285
Qy 540 GGGATGGGTTTCTTGGAAATCTCTAGAGAGCTCTCTCTGCGATGCTGAGCTCGG 599
Db 284 GGGATGGGTTTCTTGGAAATCTCTAGAGAGCTCTCTCTGCGATGCTGAGCTCGG 225
Qy 600 CAGAGAGCCCGAGTGTCTTCTCGGATGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 659
Db 224 CAGAGAGCCCGAGTGTCTTCTCGGATGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 165
Qy 660 GCTTATGTGAATTCATGCTCTTCTTCTCATCAGAGAGTATGGAATGCTTCTTCT 719
Db 164 GCTTATGTGAATTCATGCTCTTCTTCTCATCAGAGAGTATGGAATGCTTCTTCT 105
Qy 720 TTTGTTGTCTGATTTATGTTTATTTTATGATTAACAAAGTTTATTTAGCATTTCTG 779
Db 104 TTTGTTGTCTGATTTATGTTTATTTTATGATTAACAAARWTTTATTTAGCATTTCTG 45
Qy 780 AAAGAGGAAAGTAAATGTACAACTTAAATAAAG 817
Db 44 AAAGAGGAAAGTAAATGTACAACTTAAATAAAG 7

RESULT 3

AL548687/c 1001 bp mRNA linear EST 16-FEB-2001

LOCUS AL548687.LPI.NFL006.PL2 Homo sapiens cDNA clone GS0D1036Y13 3

DEFINITION Prime, mRNA sequence.

ACCESSION AL548687

VERSION AL548687.1 GI:12883941

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRI cedex - France

Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr.

Location/Qualifiers

1. 1001

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="GS0D1036Y13"

/clone_lib="LPI.NFL006.PL2"

/issue_type="placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime and

enriched, double-stranded cDNA was digested with NotI and

cloned into the NotI and EcoRV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact: Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax: (1) 301 610 8371

Email: fliang@life.com URL: http://fulllength.invitrogen.com"

BASE COUNT 265 a 279 c 237 g 217 t 3 others

ORIGIN

Query Match 89.9%; Score 785; DB 9; Length 1001;

Best Local Similarity 99.8%; Freq. No. 2, 2e-133;

Matches 807; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

1 CTCAGGAGATGTTCATGAGATATGACCGAGAGCTGAGCTGCGCTTG 60

Db 809 CTCAGGAGATGTTCATGAGATATGAGAACTAGACCGCAACGACATGACTGGGCTTG 750
Qy 61 CCGTGAATCCCTCCACGCTGATCTTGAAGTGAAGAGAACTCCCAATTAATTCAT 120
Db 749 CCGTGAATCCCTCCACGCTGATCTTGAAGTGAAGAGAACTCCCAATTAATTCAT 630
Qy 121 CTATGAGAGCTGCCGGGCAATTAAGAACAGTACCGCTGAGAGAGCTGATCTCCG 180
Db 689 CTATGAGAGCTGCCGGGCAATTAAGAACAGTACCGCTGAGAGAGCTGATCTCCG 631
Qy 181 CTGCTCCGACAGAGAGAAATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 240
Db 630 CTGCTCCGACAGAGAGAAATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 571
Qy 241 GGGGCTGTGTATGATGATGATCTCTTCTGAGAGCTCCAGAGTCTACCTGATCCG 300
Db 570 GGGGCTGTGTATGATGATGATCTCTTCTGAGAGCTCCAGAGTCTACCTGATCCG 511
Qy 301 GGTGACAGAGAGAACAGAGAGCTGCCCTGCGACCGCTGAGAGCTCCGAGATGACAA 360
Db 510 GGTGACAGAGAGAACAGAGAGCTGCCCTGCGACCGCTGAGAGCTCCGAGATGACAA 451
Qy 361 GAGAGAGCTGTGAGAGACATATGCTCTGTGACCGCTCTGCGCAAGAGACT 636 419
Db 450 GAGAGAGCTGTGAGAGACATATGCTCTGTGACCGCTCTGCGCAAGAGACT 636 391
Qy 420 AAGGAGGAGAGACTATGTGTAGCTTTTAAATAGAGAGTGAATGATCGGATTTAGT 479
Db 390 AAGGAGGAGAGACTATGTGTAGCTTTTAAATAGAGAGTGAATGATCGGATTTAGT 331
Qy 480 GATCATTAGGAGCTGAGAGCTGTCTTCTGAGAGAGTGAAGAGCTGCTTCTGAGCA 539
Db 330 GATCATTAGGAGCTGAGAGCTGTCTTCTGAGAGAGTGAAGAGCTGCTTCTGAGCA 271
Qy 540 GGGATGGGTTTCTTGGAAATCTCTAGAGAGCTCTCTCTGCAATGCTGAGCTGCG 599
Db 270 GGGATGGGTTTCTTGGAAATCTCTAGAGAGCTCTCTCTGCAATGCTGAGCTGCG 211
Qy 600 CAGAGAGCCCGAGTGTCTTCTCGGATGATTTCTTCTTCTTCTTCTTCTTCTTCTT 659
Db 210 CAGAGAGCCCGAGTGTCTTCTCGGATGATTTCTTCTTCTTCTTCTTCTTCTTCTT 151
Qy 660 GCTTATGTGAATTCATGCTCTTCTTCTCATCAGAGAGTATGGAATGCTTCTTCT 719
Db 150 GCTTATGTGAATTCATGCTCTTCTTCTCATCAGAGAGTATGGAATGCTTCTTCT 91
Qy 720 TTTGTTGTCTGATTTATGTTTATTTTATGATTAACAAAGTTTATTTAGCATTTCTG 779
Db 90 TTTGTTGTCTGATTTATGTTTATTTTATGATTAACAAAGTTTATTTAGCATTTCTG 31
Qy 780 AAAGAGGAAAGTAAATGTACAACTTAA 808
Db 30 AAAGAGGAAAGTAAATGTACAACTTAA 2

RESULT 4

BE621676/c 923 bp mRNA linear EST 20-OCT-2000

LOCUS BE621676 60149368311 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895701 3'

DEFINITION mRNA sequence.

ACCESSION BE621676

VERSION BE621676.1 GI:9892714

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 923)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lim.gov>

Plate: LLM9687 row: f column: 22

High quality sequence start: 12

High quality sequence stop: 771.

FEATURES
 Location/Qualifiers

1..923

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3895701"

/clone_lib="NIH MGC 70"

/tissue_type="epithelioid carcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."

BASE COUNT 244 a 258 c 220 g 201 t

ORIGIN

Query Match 89.7%; Score 783.4; DB 10; Length 923;

Best Local Similarity 98.8%; Pred. No. 4,5e-133;

Matches 831; Conservative 0; Mismatches 6; Indels 4; Gaps 4;

QY 1 CTCACGATATGTTCAACTGTGAAGATATCTGACCGCAACGAGTCACTGGGCTT 59

DB 843 CTCACGAGAAATGTCAACTGTGAAGAA-ACTGACCGCAACGAGTCACTGGGCTT 785

QY 60 GCGTGCACTCTCCACGCTGCTTGAAGTGAAGAACTCTGAGAGAGGCTGATCTCC 119

DB 784 GCGTGCACTCTCCACGCTGCTTGAAGTGAAGAACTCTGAGAGAGGCTGATCTCC 725

QY 120 TCTATGAGAGCTGCGGGGCAATAGAACAGCTACCGCTCTGAGAGAGGCTGATCTCC 179

DB 724 TCTATGAGAGCTGCGGGGCAATAGAACAGCTACCGCTCTGAGAGAGGCTGATCTCC 665

QY 180 GCTGCTTCGCGCAGAGAGAAATCTCCCTGCGCTTGGCTCAAAAGTGGTGTCTGG 239

DB 664 GCTGCTTCGCGCAGAGAGAAATCTCCCTGCGCTTGGCTCAAAAGTGGTGTCTGG 605

QY 240 CGGGGCTGTTCGATGATGATCTCTCTGAGAGGCTCCCATGCTGATCTATCC 299

DB 604 CGGGGCTGTTCGATGATGATCTCTCTGAGAGGCTCCCATGCTGATCTATCC 545

QY 300 GGTGGCAGCGAGAAACAGAGAGCTGCGCAGCTGAGCTCCGAGATGACA 359

DB 544 GGTGGCAGCGAGAAACAGAGAGCTGCGCAGCTGAGCTCCGAGATGACA 485

QY 360 AGAGAGAGCTGTGAAGAACATATATCTGTGACCGCTCTGCGCAAGAGACT-GG 418

DB 484 AGAGAGAGCTGTGAAGAACATATATCTGTGACCGCTCTGCGCAAGAGACT-GG 425

QY 419 GAAGAGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGGATGATCGATTTGAG 478

DB 424 GAAGAGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGGATGATCGATTTGAG 365

QY 479 TGATCATTAAGGCTAGGCTGTTTCTCTGGAGTAGAGACGGCTGCTTCTGCTGGC 538

DB 364 TGATCATTAAGGCTAGGCTGTTTCTCTGGAGTAGAGACGGCTGCTTCTGCTGGC 305

QY 539 AGGATAGGCTTTGTTGAAATCTCTAGAGAGCTCTCTCTGATGCTGAGCTG 598

DB 304 AGGATAGGCTTTGTTGAAATCTCTAGAGAGCTCTCTCTGATGCTGAGCTG 245

QY 599 GCGAGAGCCCGAGTGTCTCTGAGATGATCTTCTCTCGAGTGAAGTTCCTT 658

DB 244 GCGAGAGCCCGAGTGTCTCTGAGATGATCTTCTCTCGAGTGAAGTTCCTT 185

QY 659 TGCTTATGTTAATTCATTCGCTTTTCTCATCATCAGAGATGATGTGGATCGTTTC 718

DB 184 TGCTTATGTTAATTCATTCGCTTTTCTCATCATCAGAGATGATGTGGATCGTTTC 125

QY 719 TTTTGTTCCTGATTTATGTTTATTAAGTAAACAAAGTTTATTTATGATCTC 778

DB 124 TTTTGTTCCTGATTTATGTTTATTAAGTAAACAAAGTTTATTTATGATCTC 65

QY 779 GAAAGAGAGAAATGTAATGTACAGTTTAAATAAAAGGAGGCTTCCCTTGAATAA 838

DB 64 GAAAGAGAGAAATGTAATGTACAGTTTAAATAAAAGGAGGCTTCCCTTGAATAA 6

QY 839 T 839

DB 5 T 5

RESULT 5

AL575665/c AL575665 1061 bp mRNA linear EST 16-FEB-2001

LOCUS AL575665 LTI_NFL006_PU2 Homo sapiens cDNA clone CSOD1069Y17 3

DEFINITION prime, mRNA sequence.

ACCESSION AL575665

VERSION AL575665.1 GI:12937050

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1061)

AUTHORS Li,W.B., Gruber,C., Jesse,J., and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..1061

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSOD1069Y17"

/clone_lib="LTI_NFL006_PU2"

/tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>"

BASE COUNT 281 a 301 c 250 g 226 t 3 others

ORIGIN

Query Match 89.6%; Score 782.4; DB 9; Length 1061;

Best Local Similarity 99.5%; Pred. No. 6,4e-133;

Matches 805; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 CTCACGATATGTTCAACTGTGAAGATATCTGACCGCAACGAGTCACTGGGCTTG 60

DB 808 CTCACGATATGTTCAACTGTGAAGATATCTGACCGCAACGAGTCACTGGGCTTG 749

QY 61 CCGTGATCTTCCACGCTGTGATCTTGAAGAGAACTCTGCAATTAATTCAAT 120

DB 748 CCGTGATCTTCCACGCTGTGATCTTGAAGAGAACTCTGCAATTAATTCAAT 689

QY 121 CTATGAGAGCTGCGGGGCAATAGAAACAGCTACCGCTCTGAGAGGCTGATCTCCG 180

DB 688 CTATGAGAGCTGCGGGGCAATAGAAACAG-TCACGCTCTGAGAGGCTGCAATTCGCG 630

```

QY 181 CTGCTCCGCCACAGAGAAATCCCTCCCTGCTCAAGTGTGTCTGAC 240
Db 629 CTGCTCCGCCACAGAGAAATCCCTCCCTGCTCAAGTGTGTCTGAC 570
QY 241 GGGGCTGTGTGATGCTGTATCTCTTCTTGGAGGCTCCATGCTCACTGATCCG 300
Db 569 GGGGCTGTGTGATGCTGTATCTCTTCTTGGAGGCTCCATGCTCACTGATCCG 510
QY 301 GATGACACGAGGAAACAGAGAGGCTGCTGCGACGCTGTGAGTCCGAGATGACAA 360
Db 509 GATGACACGAGGAAACAGAGAGGCTGCTGCGACGCTGTGAGTCCGAGATGACAA 450
QY 361 GAGACAGCTGTGTAAGAACATATGTCCTGTACCCCTGCTGCGCAAGAGACT- GGG 419
Db 449 GAGACAGCTGTGTAAGAACATATGTCCTGTACCCCTGCTGCGCAAGAGACTGGG 390
QY 420 AAGGAGGAGGAACTATGTGTAGCTTTTAAATAGAGGATGACTCGGATTTAGT 479
Db 389 AAGGAGGAGGAACTATGTGTAGCTTTTAAATAGAGGATGACTCGGATTTAGT 330
QY 480 GATCATTAAGGCTGAGGCTGTCTCTCTGAGAGTGAAGAGGCTGCTGCTGCA 539
Db 329 GATCATTAAGGCTGAGGCTGTCTCTCTGAGAGTGAAGAGGCTGCTGCTGCA 270
QY 540 GGAATGAGTTTGTCTTGAAGATCTCTAGAGAGCTCCTGCTGCAATGGCTGAGTCTGG 599
Db 269 GGAATGAGTTTGTCTTGAAGATCTCTAGAGAGCTCCTGCTGCAATGGCTGAGTCTGG 210
QY 600 CAGAGAGCCCGAGTTGTTCTCGCTGATGATTTCTTCTTCAAGTAGATTTTCTT 659
Db 209 CAGAGAGCCCGAGTTGTTCTCGCTGATGATTTCTTCTTCAAGTAGATTTTCTT 150
QY 660 GCTATATGTAATTCATTCCTCTTCTCATACAGAGTGAATGATGTTCT 719
Db 149 GCTATATGTAATTCATTCCTCTTCTCATACAGAGTGAATGATGTTCT 90
QY 720 TTGTGTTCTGATTAATGTTTAAATAGATTAAGATTAAGATTCG 779
Db 89 TTGTGTTCTGATTAATGTTTAAATAGATTAAGATTAAGATTCG 30
QY 780 AAGAGAGAAAGTAATGTACAGTTTA 808
Db 29 AAGAGAGAAAGTAATGTACAGTTTA 1

RESULT 6
AUS73986/c 983 bp mRNA linear EST 16-FEB-2001
LOCUS AUS73986
DEFINITION prime, mRNA sequence.
ACCESSION AUS73986
VERSION AUS73986.1 GI:12933753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 983)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1..983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1053YK07"
/clone_1b="LTI NFL006_PL2"
/risue_type="placental"
/Note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

```

```

QY 1 CTCAGGATATATGTTCACTATGAGAAATACGACCGCCAAAGCATCTGAGCCCTTG 60
Db 811 CTCAGGATATATGTTCACTATGAGAAATACGACCGCCAAAGCATCTGAGCCCTTG 752
QY 61 CCGTGATTCCTTCCACGCTGATCTTTGACGTGAGAGAACTCCTGCATTAATTCTAT 120
Db 751 CCGTGATTCCTTCCACGCTGATCTTTGACGTGAGAGAACTCCTGCATTAATTCTAT 692
QY 121 CTATGAGAGGCTGCGGGGCAATTAAGAA-CAGCTAACGCTCTGAGAGGCTGCAATGCTCC 179
Db 691 CTATGAGAGGCTGCGGGGCAATTAAGAACTGAGAGGCTGCAATGCTCC 632
QY 180 GCTGCTTCCGACAGAGAAATCCTCCCTGAGCCCTTGGCTCAAGAGGATGCTGG 239
Db 631 GCTGCTTCCGACAGAGAAATCCTCCCTGAGCCCTTGGCTCAAGAGGATGCTGG 572
QY 240 CGGGAGCTGTTGATGATGATGATCTCTTCTGAGAGCTTCATGATCTGATTC 299
Db 571 CGGGAGCTGTTGATGATGATGATCTCTTCTGAGAGCTTCATGATCTGATTC 512
QY 300 GGGTGGACAGAGAAACAGAGAGCGTGCCTGCGCAACCTGCGAGCTCCGAGATGACA 359
Db 511 GGGTGGACAGAGAAACAGAGAGCGTGCCTGCGCAACCTGCGAGCTCCGAGATGACA 452
QY 360 AGAGACAGCTGTGTAAGAACATATGTCCTGTGACCGCCTGCGCAAGAGACT- GG 418
Db 451 AGAGACAGCTGTGTAAGAACATATGTCCTGTGACCGCCTGCGCAAGAGACTGGG 392
QY 419 GAAGGAGAGGAGAACTATGTGTAGCTTTTAAATAGAGGATGACTCGGATTTGAG 478
Db 391 GAAGGAGAGGAGAACTATGTGTAGCTTTTAAATAGAGGATGACTCGGATTTGAG 332
QY 479 TGATCATTAAGGCTGAGGCTGTCTTCTCTGAGAGTGAAGAGGCTGCTTCTGCTGAGC 538
Db 331 TGATCATTAAGGCTGAGGCTGTCTTCTCTGAGAGTGAAGAGGCTGCTTCTGCTGAGC 272
QY 539 AAGGATGAGTTTGTCTTGAAGAACTCTAGAGGCTCCTCTGCAATGAGCTGAGCTG 598
Db 271 AAGGATGAGTTTGTCTTGAAGAACTCTAGAGGCTCCTCTGCAATGAGCTGAGCTG 212
QY 599 GCAGAGCCCCGAGTTGTTCTCTGCTGATGATTTCTTCTGCTGCTGCTGCTGCTGCT 658
Db 211 GCRGAGCCCCGAGTTGTTCTCTGCTGATGATTTCTTCTGCTGCTGCTGCTGCTGCT 152
QY 659 TGCTTAATGTAATTCATTCCTCTTCTTCTCATACAGAGTGAATGATGATGCTTC 718
Db 151 TGCTTAATGTAATTCATTCCTCTTCTTCTCATACAGAGTGAATGATGATGCTTC 92
QY 719 TTGTGTTCTGATTAATGTTTAAATAGATTAAGATTAAGATTCG 778
Db 91 TTGTGTTCTGATTAATGTTTAAATAGATTAAGATTAAGATTCG 32
QY 779 GAAGAGAGAAAGTAATGTACAGTTTA 808
Db 31 GAAGAGAGAAAGTAATGTACAGTTTA 2

```

BASE COUNT 258 a 276 c 230 g 212 t 7 others
 ORIGIN

Query Match 89.5%; Score 781.6; DB 9; Length 983;
 Best Local Similarity 99.1%; Pred. No. 9.3e-133;
 Matches 803; Conservative 3; Mismatches 2; Indels 2; Gaps 2;

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetechn.com URL : http://fulllength.invitrogen.com"

```

RESULT 7
AL575631/c 938 bp mRNA linear EST 16-FEB-2001
LOCUS AL575631 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0D1069Y006 3
DEFINITION prime, mRNA sequence.
ACCESSION AL575631 GI:12936982
VERSION 1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 938)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006-EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1069Y006"
/clone_lib="LTI.NFL006.PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT)-primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 250 a 260 c 223 g 201 t 4 others
ORIGIN
Query Match 88.9%; Score 776.2; DB 9; Length 938;
Best Local Similarity 99.5%; Pred. No. 9,2e-132;
Matches 797; Conservative 2; Mismatches 0; Indels 2; Gaps 2;

QY 2 TCACGATATGTTCACTATGAGAAATGACACCGCAACGAGTCACTGAGCCCTTGC 61
Db 800 TCACGATATGTTCACTATGAGAAATGACACCGCAACGAGTCACTGAGCCCTTGC 741
QY 62 CGTGCATCCTTCCACGCTGGTACTTGAAGTGAAGAGAACTCTGCATTAATCTTCATC 121
Db 740 CGTGCATCCTTCCACGCTGGTACTTGAAGTGAAGAGAACTCTGCATTAATCTTCATC 681
QY 122 TATGAGAGGTGCTGGCGGCAATTAAGAGAGTACCTGCTGAGAGGCTTGCATGCTCCG 181
Db 680 TATGAGAGGTGCTGGCGGCAATTAAGAGAGTACCTGCTGAGAGGCTTGCATGCTCCG 622
QY 182 TGGCTTCGCGCAGCAGAGATCCTCCCTGCGCTGAGAGTGAAGTGGTGGTGGGCG 241
Db 621 TGGCTTCGCGCAGCAGAGATCCTCCCTGCGCTGAGAGTGAAGTGGTGGTGGGCG 562
QY 242 GGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
Db 561 GGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 502
QY 302 GTGGCAGCGAGAGAACACAGAGAGTGGCTTGGAGACCTGTTGAGAGTCCGAGATGCAAG 361
Db 501 GTGGCAGCGAGAGAACACAGAGAGTGGCTTGGAGACCTGTTGAGAGTCCGAGATGCAAG 442
QY 362 GAGCAGCTGTGTAAGACATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 420
Db 441 GAGCAGCTGTGTAAGACATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 382
QY 421 AGGAGAGGAGACTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 480

```

```

Db 381 AGGAGAGGAGACTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 322
QY 481 ATCATTAAGAGGCTGAGGTCTGTTCTCTGAGAGTGAAGACGAGCTCTCTGTTGAGCAG 540
Db 321 ATCATTAAGAGGCTGAGGTCTGTTCTCTGAGAGTGAAGACGAGCTCTCTGTTGAGCAG 262
QY 541 GGATGGGTTTGTCTTGAATCTCTGAGAGGCTCTCTGAGAGTGGCTCGAGTCTGAGC 600
Db 261 GGATGGGTTTGTCTTGAATCTCTGAGAGGCTCTCTGAGAGTGGCTCGAGTCTGAGC 202
QY 601 AGAAGCCCGAGTGTGTTCTGCGCTGATGATGATGATGATGATGATGATGATGATGATG 660
Db 201 AGAAGCCCGAGTGTGTTCTGCGCTGATGATGATGATGATGATGATGATGATGATGATG 142
QY 661 CTAAATGTAATTCATGCTCTTCTTCATCAGAGAGTGAATGTTGTTCTT 720
Db 141 CTAAATGTAATTCATGCTCTTCTTCATCAGAGAGTGAATGTTGTTCTT 82
QY 721 TTGTTGTTGATTTATGTTTATGTTTATGATTAACAAAGTTTATTAAGATTTGTA 780
Db 81 TTGTTGTTGATTTATGTTTATGTTTATGATTAACAAAGTTTATTAAGATTTGTA 22
QY 781 AAGAGGAAAGTAAATGTTAC 801
Db 21 AAGAGGAAAGTAAATGTTAC 1

RESULT 8
AL573346/c 1052 bp mRNA linear EST 16-FEB-2001
LOCUS AL573346 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0D1042YC20 3
DEFINITION prime, mRNA sequence.
ACCESSION AL573346
VERSION AL573346.1 GI:12932499
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1052)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006-EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1042YC20"
/clone_lib="LTI.NFL006.PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT)-primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 279 a 298 c 248 g 222 t 5 others
ORIGIN
Query Match 88.9%; Score 776; DB 9; Length 1052;
Best Local Similarity 99.5%; Pred. No. 9,4e-132;
Matches 798; Conservative 1; Mismatches 1; Indels 2; Gaps 2;

QY 1 CTCACGATATGTTCACTATGAGAAATGACACCGCAACGAGTCACTGAGCCCTTGC 60

```

```

Db      802 CTCACGCATATGTTCAACTATGAAGTAATACGACCGCCACGACAGTCACTGGGCTTG 743
Qy      61 CCGTCATCTCTTCCACGCTGACTTTGAGGTGAGAGAACTCTCTGAAATTAATTCAT 120
Db      742 CCGTCATCTCTTCCACGCTGACTTTGAGGTGAGAGAACTCTCTGAAATTAATTCAT 683
Qy      121 CTATGAGAGCTGCCGGGCAATAG-AAACGCTACCGCTCTGAGAGAGCTCATGCTCC 179
Db      682 CTATGAGAGCTGCCGGGCAATAGAAACACTACCGCTCTGAGAGAGCTCATGCTCC 623
Qy      180 GCTGCTTCCGACGACGAGAGAAATCTCCCTCCCTGAGGCTCAAAAGTGAGTTCG 239
Db      622 GCTGCTTCCGACGACGAGAGAAATCTCCCTCCCTGAGGCTCAAAAGTGAGTTCG 563
Qy      240 CGGGGCTGCTGAGAGAGTGTGATCTCTCTGAGAGGCTCAAGTGTGATTC 299
Db      562 CGGGGCTGCTGAGAGAGTGTGATCTCTCTGAGAGGCTCAAGTGTGATTC 503
Qy      300 GGGTGGCAAGGAGGAGACGAGAGGCTGCGACCGTCTGAGAGCTCCGAGATGACA 359
Db      502 GGGTGGCAAGGAGGAGACGAGAGGCTGCGACCGTCTGAGAGCTCCGAGATGACA 443
Qy      360 AGGAGACGCTGTGAAGAACACATATGCTGTGACCGCCCTGCGCAAGAGACT-GG 418
Db      442 AGGAGACGCTGTGAAGAACACATATGCTGTGACCGCCCTGCGCAAGAGACTGGG 383
Qy      419 GAAAGGAGGAGACATATGAGTGTGAGCTTTTAAATAGAGGATGACTCGAATTTGAG 478
Db      382 GAAAGGAGGAGACATATGAGTGTGAGCTTTTAAATAGAGGATGACTCGAATTTGAG 323
Qy      479 TGATCATTAAGGAGCTGAGAGTGTGTTCTCTGAGAGTGAAGAGGCTCTCTGAGTGGC 538
Db      322 TGATCATTAAGGAGCTGAGAGTGTGTTCTCTGAGAGTGAAGAGGCTCTCTGAGTGGC 263
Qy      539 AAGGATGAGTGTGCTTGAAGAAATCCTCTAGAGAGGCTCTCTGAGAGGCTCTGAGTGG 598
Db      262 AAGGATGAGTGTGCTTGAAGAAATCCTCTAGAGAGGCTCTCTGAGAGGCTCTGAGTGG 203
Qy      599 GCAGAGCCCGAGGTTGTTCTCGCTATGCAATTTCTTCTCCAGGTAGAGTTTCTT 658
Db      202 GCAGAGCCCGAGGTTGTTCTCGCTATGCAATTTCTTCTCCAGGTAGAGTTTCTT 143
Qy      659 TGCTATGTTGAATTCATGCTCTTCTTCTCAACAGAGAGATGTTGAATGCTTC 718
Db      142 TGCTATGTTGAATTCATGCTCTTCTTCTCAACAGAGAGATGTTGAATGCTTC 83
Qy      719 TTTGTTGCTGATTTAATGTTTAAATAGATTAACAAAGTTTAAATAGATTC 778
Db      82 TTTGTTGCTGATTTAATGTTTAAATAGATTAACAAAGTTTAAATAGATTC 23
Qy      779 GAAAGAGGAAAGTAAATGTA 800
Db      22 GAAAGAGGAAAGTAAATGTA 1

RESULT 9
AUS76440/c 877 bp mRNA linear EST 16-FEB-2001
LOCUS AUS76440 LTI NFLO06.PL2 Homo sapiens cDNA clone CS001075Yc08.3
DEFINITION prime, mRNA sequence.
ACCESSION AUS76440
VERSION AUS76440.1 GI:12938586
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eumetazoa; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 877)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

```

```

FEATURES
source
  1..877
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="CS001075Yc08"
  /tissue_type="Placenta"
  /note="Vector: pCMVSPORT 6; Site 1: NotI, 1st strand cDNA
  was primed with a NotI-oligo(dT) primer. Five prime end
  enriched, double-stranded cDNA was digested with Not I and
  cloned into the Not I and Eco RV sites of the pCMVSPORT 6
  vector. Library was normalized. Library was constructed by
  Life Technologies. Contact : Feng Liang Life Technologies,
  Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
  Email : fliang@lifetech.com URL :
  http://fulllength.invitrogen.com"

BASE COUNT 226 a 245 c 205 g 192 t 9 others
ORIGIN

Query Match      88.3%; Score 770.8; Ds 9; Length 877;
Best Local Similarity 98.1%; Pred. No. 9,1e-131;
Matches 783; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Qy      1 CTCACGCATATGTTCAACTATGAAGTAATACGACCGCCACGACAGTCACTGGGCTTG 60
Db      809 CTCACGCATATGTTCAACTATGAAGTAATACGACCGCCACGACAGTCACTGGGCTTG 750
Qy      61 CCGTCATCTCTTCCACGCTGACTTTGAGGTGAGAGAACTCTCTGAAATTAATTCAT 120
Db      749 CCGTCATCTCTTCCACGCTGACTTTGAGGTGAGAGAACTCTCTGAAATTAATTCAT 690
Qy      121 CTATGAGAGCTGCCGGGCAATAGAAACAGTACCGCTCTGAGAGGCTCATGCTCCG 180
Db      682 CTATGAGAGCTGCCGGGCAATAGAAACAGTACCGCTCTGAGAGGCTCATGCTCCG 630
Qy      181 CTGCTTCCGACGAGAGAGAAATCCTCCCTGCCCCCTTGAGCTCAAAAGTGTGTTCTG 240
Db      622 CMGCTTCCGACGAGAGAGAAATCCTCCCMGCCCTTGCTCAAAAGTGTGATCTGCG 570
Qy      241 GGGGCTGTGTGATGTGTGATCTCTCTGAGAGCTCCATGCTTACTGATCCG 300
Db      569 GGGGCTGTGTGATGTGTGATCTCTCTGAGAGCTCCATGCTTACTGATCCG 510
Qy      301 GGTGGACGAGAGAACACAGAGACCGTCCCGACCGTCTGAGCTCCGAGATGACAA 360
Db      509 GGTGGACGAGAGAACACAGAGACCGTCCCGACCGTCTGAGCTCCGAGATGACAA 450
Qy      361 GAGAGAGCTGTGAGAGACACATATGCTGTGACCGCCCTGCGCAAGAGACT-GGG 419
Db      449 GAGAGAGCTGTGAGAGACACATATGCTGTGACCGCCCTGCGCAAGAGACTGGGG 390
Qy      420 AAGGAGGAGAGACATATGCTGTGAGCTTTTAAATAGAGAGTACTCGAATTTGAGT 479
Db      389 AAGGAGGAGAGACATATGCTGTGAGCTTTTAAATAGAGAGTACTCGAATTTGAGT 330
Qy      480 GATCATTAAGGCTGAGAGTCTGTTTCTCTGAGAGGTAGAGAGGCTGTTCTGCTGAGCA 539
Db      329 GATCATTAAGGCTGAGAGTCTGTTTCTCTGAGAGGTAGAGAGGCTGTTCTGCTGAGCA 270
Qy      540 GGGATGGGTTTGGTGAATCTCTGAGAGGCTCTCTGCGATGAGCTGAGCTGG 599
Db      269 GGGATGGGTTTGGTGAATCTCTGAGAGGCTCTCTGCGATGAGCTGAGCTGG 210
Qy      600 CAGACGCCCGAGTGTTCCTCGCTGATGATTTCTTCTCCAGAGTATGATTTCTT 659
Db      209 CAGACGCCCGAGTGTTCCTCGCTGATGATTTCTTCTCCAGAGTATGATTTCTT 150
Qy      660 GCTATGTAATTCATGCTCTTTTCTATACAGAGAGTGTGAGATGTTCT 719

```

```
Db 149 GCTTAGTGAATTCATTCGATCTTTCTCATCAGAGAGTAGATTGGATCGTGTCT 90
Qy 720 TTGTTGTCGATTTATGATTTTATTAAGTATTAACAAAGATTTTATTAACATTCG 779
Db 89 TTGTTGTCGATTTATGATTTTATTAAGTATTAACAAAGATTTTATTAACATTCG 30
Qy 780 AAAGAGAGAAAGTAAAT 797
Db 29 AAAGAGAGAAAGTAAAT 12

RESULT 10
AL576831/c 950 bp mRNA linear EST 16-FEB-2001
LOCUS AL576831 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1079B07 3
DEFINITION prime, mRNA sequence.
ACCESSION AL576831
VERSION AL576831.1 GI:12939362
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 950)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..950
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1079B07"
/clone_id="LTI_NFL006_P12"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 252 a 263 c 224 g 208 t 3 others
ORIGIN
Query Match 88.3%; Score 770.6; DB 9; Length 950;
Best Local Similarity 99.3%; Pred. No. 9, 5e-13;
Matches 803; Conservative 2; Mismatches 1; Indels 3; Gaps 3;

Qy 1 CTCAGGATATGTTCAACTAAGAGATCTGACCGGCAAGCATGCTGAGCCTTG 60
Db 808 CTCAGGATATGTTCAACTAAGAGATCTGACCGGCAAGCATGCTGAGCCTTG 749
Qy 61 CCGTGATCTCTCCACGCTGTACTTTGACGTGGAGAGAACTCTGCAATTAATTCA 120
Db 748 CCGTGATCTCTCCACGCTGTACTTTGACGTGGAGAGAACTCTGCAATTAATTCA 689
Qy 121 CTATGAGGCTGCGGGGCAATTAAGACACTACCGCTGAGAGGCTGCAATGCTCCG 180
Db 688 CTATGAGGCTGCGGGGCAATTAAGACAG-TACCGCTGAGAGGCTGCAAT-CTCCG 631
Qy 181 CTGCTTCGCGCAGAGAGAAATCTCCCTGCTGCTGCTCAAGAGTGTGTTCTGCG 240
Db 630 CTGCTTCGCGCAGAGAGAAATCTCCCTGCTGCTGCTCAAGAGTGTGTTCTGCG 571
Qy 241 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300

Db 570 GGGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 511
Qy 301 GGTGGCAGCAGAGAAACAGAGAGCGTCCGCGCACCGTCTGAGCTCGGAGATGACAA 360
Db 510 GGTGGCAGCAGAGAAACAGAGAGCGTCCGCGCACCGTCTGAGCTCGGAGATGACAA 451
Qy 361 GAGCAGCTGTGAGAGAAACATATGCTCTGAGACCGCCCTGTGCTCAAGAGACT-GGG 419
Db 450 GAGCAGCTGTGAGAGAAACATATGCTCTGAGACCGCCCTGTGCTCAAGAGACTG 391
Qy 420 AAGGAGGGGAGACATATGAGCTTTTATTAAGAGGATTAAGTCTGAGATTGAGT 479
Db 390 AAGGAGGGGAGACATATGAGCTTTTATTAAGAGGATTAAGTCTGAGATTGAGT 331
Qy 480 GATCATTAGAGCTGAGAGCTGTGTTCTCTGAGAGGTGAGACGCTGCTCTGCTGAGCA 539
Db 330 GATCATTAGAGCTGAGAGCTGTGTTCTCTGAGAGGTGAGACGCTGCTCTGCTGAGCA 271
Qy 540 GGGATGGGTTTCTTTGGAATACTCTGAGAGGCTCTCTGAGATGGCTGAGATCTGG 599
Db 270 GGGATGGGTTTCTTTGGAATACTCTGAGAGGCTCTCTGAGATGGCTGAGATCTGG 211
Qy 600 CAGAGCCCGAGATTGTTCTCTGCTGATGATGATGATGATGATGATGATGATGATG 659
Db 210 CAGAGCCCGAGATTGTTCTCTGCTGATGATGATGATGATGATGATGATGATGATG 151
Qy 660 GCTTAGTTGAATTCATGCTCTTTTCTCATCAGAGAGTATGTTGATGATGATGAT 719
Db 150 GCTTAGTTGAATTCATGCTCTTTTCTCATCAGAGAGTATGTTGATGATGATGAT 91
Qy 720 TTGTTGTCGATTTATGATTTTATTAAGTATTAACAAAGTTTTATTAAGATTCG 779
Db 90 TTGTTGTCGATTTATGATTTTATTAAGTATTAACAAAGTTTTATTAAGATTCG 31
Qy 780 AAAGAGAGAAAGTAAATGTAAGTTTA 808
Db 30 AAAGAGAGAAAGTAAATGTAAGTTTA 2

RESULT 11
AL575764/c 1062 bp mRNA linear EST 16-FEB-2001
LOCUS AL575764 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1070Y15 3
DEFINITION prime, mRNA sequence.
ACCESSION AL575764
VERSION AL575764.1 GI:12937249
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1062)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..1062
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1070Y15"
/clone_id="LTI_NFL006_P12"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
```

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : filang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 277 a 300 c 252 g 227 t 6 others
 ORIGIN

Query Match 86.2%; Score 752.8; DB 9; Length 1062;
 Best Local Similarity 98.5%; Pred. No. 1.2e-127;
 Matches 786; Conservative 5; Mismatches 4; Indels 3; Gaps 3;

```

QY 1 CTCGAGGATATGTTCACTAATGAAGATATCTGACACGCGCCAGCGCATCTGCGGCTTG 60
DB CTCGAGGATATGTTCACTAATGAAGATATCTGACACGCGCCAGCGCATCTGCGGCTTG 745
QY 61 CCGTCATCTCTCCACGCTGACTTTGAGCTGGAGAGAACTCTGCAATATCTTAT 120
DB 744 CCGTCATCTCTCCACGCTGACTTTGAGCTGGAGAGAACTCTGCAATATCTTAT 685
QY 121 CTATGAGAGCTGCGGCGCATTAAGAACAGCTACCGCTTGAAGAGGCTGCTGCTCCG 180
DB 684 CTATGAGAGCTGCGGCGCATTAAGAACAGCTACCGCTTGAAGAGGCTGCTGCTCCG 627
QY 181 CTGCTCCGCGCAGAGAGAAATCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 628 CTGCTCCGCGCAGAGAGAAATCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 567
QY 241 GGGGCTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 566 GGGGCTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 507
QY 301 GGTGACAGAGAAACAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 506 GGTGACAGAGAAACAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 447
QY 361 GAGAGAGTGTGAAGAACATATGCTGTGACCGCTTGTGCGCAAGAGACT-GGG 419
DB 446 GAGAGAGTGTGAAGAACATATGCTGTGACCGCTTGTGCGCAAGAGACTGAGG 387
QY 420 AAGGAGGAGGAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 479
DB 386 AAGGAGGAGGAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 327
QY 480 GATCATTAAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539
DB 326 GATCATTAAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 267
QY 540 GGGATGGGTTGCTTTGGAATCTCTAGAGAGGCTCTCTGCAATGAGCTGCACTGG 599
DB 266 GGGATGGGTTGCTTTGGAATCTCTAGAGAGGCTCTCTGCAATGAGCTGCACTGG 207
QY 600 CAGCAGCCCGGAGTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
DB 206 CAGCAGCCCGGAGTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 147
QY 660 GCTTATGTTGAATTCATGCTCTTCTCATCAGAGAGTATGTTGAATGTTCT 719
DB 146 GCTTATGTTGAATTCATGCTCTTCTCATCAGAGAGTATGTTGAATGTTCT 87
QY 720 TTTGTTGCTGATTTATGTTTATTAAGTATTAACAAAGTTTATTAAGATTTCTG 779
DB 86 TTTGTTGCTGATTTATGTTTATTAAGTATTAACAAAGTTTATTAAGATTTCTG 27
QY 780 AAAAGAGAGAGTAAAT 797
DB 26 AAAAGAGAGAGTAAAT 9

```

RESULT 12
 AL560664/c
 LOCUS AL560664 964 bp mRNA linear EST 16-FEB-2001
 DEFINITION AL560664 L11 NFL010 BC2 Homo sapiens cDNA clone CS0DL003YCI1 3
 primer, mRNA sequence.
 ACCESSION AL560664

VERSION AL560664.1 GI:12907346
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 964)
 AUTHORS Li, M.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 FEATURES
 source
 1. 964
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DL003YCI17"
 /clone_1b="L11_NFL010_BC2"
 /sex="male"
 /issue_type="B cells from Burkitt lymphoma"
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 255 a 274 c 233 g 199 t 3 others
 ORIGIN

Query Match 85.7%; Score 748; DB 9; Length 964;
 Best Local Similarity 99.1%; Pred. No. 1.2e-126;
 Matches 770; Conservative 2; Mismatches 3; Indels 2; Gaps 2;

```

QY 1 CTCGAGGATATGTTCACTAATGAAGATATCTGACACGCGCCAGCGCATCTGCGGCTTG 60
DB 787 CTCGAGGATATGTTCACTAATGAAGATATCTGACACGCGCCAGCGCATCTGCGGCTTG 728
QY 61 CCGTCATCTCTCCACGCTGACTTTGAGCTGGAGAGAACTCTGCAATATCTTAT 120
DB 727 CCGTCATCTCTCCACGCTGACTTTGAGCTGGAGAGAACTCTGCAATATCTTAT 668
QY 121 CTATGAGAGCTGCGGCGCATTAAG-AAAGCTACCGCTTGAAGAGGCTGCACTGCC 179
DB 667 CTATGAGAGCTGCGGCGCATTAAGAAAGCTACCGCTTGAAGAGGCTGCACTGCC 608
QY 180 GCTGCTTCGCGCAGAGAGAAATCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 239
DB 607 GCTGCTTCGCGCAGAGAGAAATCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 548
QY 240 CCGGCTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 299
DB 547 CCGGCTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 488
QY 300 GGGTGGACAGGAGAGACAGAGAGCGTGGCGCACGCTGAGAGCTCGGAGATGACA 359
DB 487 GGGTGGACAGGAGAGACAGAGAGCTGCTGCGCACGCTGAGAGCTCGGAGATGACA 428
QY 360 AGGAGCAGCTGTGTAAGAACATATGCTGTGAGACCGCTGTGCGCAAGAGACT-GG 418
DB 427 AGGAGCAGCTGTGTAAGAACATATGCTGTGAGACCGCTGTGCGCAAGAGACTGG 368
QY 419 GAAAGGAGGAGAGACTATGTTGAGCTTTTAAATAGAGGATTTGATCGGATTTGAG 478
DB 367 GAAAGGAGGAGAGACTATGTTGAGCTTTTAAATAGAGGATTTGATCGGATTTGAG 308
QY 479 TGATATTAGGCTGAGGCTGTGTTCTCTGAGAGGTGAGAGCGCTCTCTGCTGCTGCGC 538

```

RESULT 13	
AL582333/c	
LOCUS	859 bp mRNA linear EST 16-FEB-2001
DEFINITION	AL582333 LTI_NFL010.BC2 Homo sapiens CDNA clone CS0DL006Y120 3
	prime, AL582333
ACCESSION	AL582333
VERSION	AL582333.1 GI:12950213
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segrete@genoscope.cns.fr, Web : www.genoscope.cns.fr

FEATURES	Location/Qualifiers
source	1. .859

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="CS0DL006Y20"  
/clone_1b="T71_NFL010_BC3"  
/sex="male"
```

l/tissue type="B" cells from Burkitt lymphoma",
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :

Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>

BASE COUNT	230 a	240 c	204 g	178 t	7 others
ORIGIN					

Query Match 85.7%; Score 747.8; DB 9; Length 859;

Matches 790; Conservative 5; Mismatches 2; Indels 4; Gaps 4

QY 1 CTCACGGATATGTTCAACTATGAAGATACTGCACCGGCCAACGCAGTCACCTGGCCTTG 60

Db 800 CTCACGATATGTTCACTATGAAGATACTGCACCGCCACGCACTACT-GGCCTTG 74

61 CCGTGCATCCTTCCACGCTGTA
CTTGAACGTGGAGAGGA
CTCCTGCAATACTTCAT 12

Db 741 CCGTGCATCCTTCCACCGCTGGTACTTTGACCTGGAGGAACTCCTGCATACTCAT 68

121 CTATGAGGCTGCCGGGCAATAA-GAACAGCTACCGCTCTGA-GGAGGCTGCATGCTC 178

QY	418	GGAGGGAGGGGAGACTATGCTGAGCTTTTAAATAGGATGTGACTGGATTGA	477
Db	381	GGAGGGAGGGGAGACTATGTAAGCTTTTAAATAGGAGATTTGAGATTGA	322
QY	478	GTGATCATTAAGGGCTAGGCTCTGTTCTCGGAGGTAGGAGCGGCTCTCGGCTCGG	537
Db	321	GTGATCATTAAGGGCTAGGCTCTGTTCTCGGAGGTAGGAGCGGCTCTCGGCTCGG	267
QY	538	CAGGGATGGGTGCTTTTGGAAATCTCTAGGAGGCTCTCTCGCATGCGCTCGAGTCT	597

Db 141 TTGCTATGTGATTCGATTCGCTCTTTTCATCAGAGATGATGCGATCGT 82

718 CTTTGTGTCGATTATCGTTTTTTAAGTAAACAAAGTTTTTATATAGCATTC 777

Db 81 CTTTGGTTGGCGAGTTTATGTTTTTTTAGTATTAACCAAATTTTATTAGCATTC 22
 QY 778 TGAAGAAAGCAAGTAAATG 798
 Db 21 TGAAGAAAGCAAGTAAATG 1

RESULT 14	AL571794/c	AL571794	990 bp	mRNA	linear	EST 16-FEB-2001
LOCUS		AL571794				
DEFINITION		AL571794 LTI NF1006 PL2	Homo sapiens	CDNA clone	CSDDI030YCI7.3	
ACCESSION		AL571794				
VERSION		AL571794.1	GI:12929445			
KEYWORDS		EST.				

KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens
Chordata; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

REFERENCE 1 (bases 1 to 990)

TITLE	Full-length cDNA libraries and normalization
-------	--

COMMENT Contact: Genoscope

Genoscope - Centre national de séquençage
BP 191 91006 EVRY cedex - France

FEATURES
Location/Qualifiers
email: seqet@genoscope.cns.fr, web: www.genoscope.cns.fr

```

1. .990
source /organism="Homo sapiens"

```

```
/db_xref="Eaxon:9606"  
/clone="CS0D1030YC17"
```



```

/clone_lib="LTI_NF1006_P12"
/tissue_type="placenta"
/notes="Vector: PCWSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

```

BASE COUNT      258 a      282 c      244 g      203 t      3 others
ORIGIN
Query Match      85.0%; Score 742.2; DB 9; Length 990;
Best Local Similarity 99.5%; Pred. No. 1,4e-125;
Matches 764; Conservative 1; Mismatches 0; Indels 2; Gaps 2;

```

```

QY 1 CTCGAGGATATGTTCACTATGAAGATACCTGACCGCCAAAGCAGTCACTGGGCTTG 60
DB 767 CTCGAGGATATGTTCACTATGAAGATACCTGACCGCCAAAGCAGTCACTGGGCTTG 708
QY 61 CCGTGCACTCTTCCACGCTGGTACTTTGACGTGAGAGAACTCTGCAATTAATTTCAT 120
DB 707 CCGTGCACTCTTCCACGCTGGTACTTTGACGTGAGAGAACTCTGCAATTAATTTCAT 648
QY 121 CTATGAGAGGTGCGGGGGCAATAGAA-CAGCTACCGCTCTGAGAGAGGCTGCAATGCTCC 179
DB 647 CTATGAGAGGTGCGGGGGCAATAGAA-CAGCTACCGCTCTGAGAGAGGCTGCAATGCTCC 588
QY 180 GCTGCTTCCGACGAGAGAGATCTCCCTGCGCTTGGCTCAAGAGTGGTGGTCTG 239
DB 587 GCTGCTTCCGACGAGAGAGATCTCCCTGCGCTTGGCTCAAGAGTGGTGGTCTG 528
QY 240 CGGAGCTGTTTGTGATGATGATCTCTTCTCTGAGACCTCAGTGTCTACTATCTC 299
DB 527 CGGAGCTGTTTGTGATGATGATCTCTTCTCTGAGACCTCAGTGTCTACTATCTC 468
QY 300 GGGTGGCAGGAGAGAACAGAGAGGCTGCGCACCGCTGAGACTCGGAGATGACA 359
DB 467 GGGTGGCAGGAGAGAACAGAGAGGCTGCGCACCGCTGAGACTCGGAGATGACA 408
QY 360 AGGAGCAGCTGGGAGAAACATATGTCGTGACCGCCCTGCGCAAGAGACT-GG 418
DB 407 AGGAGCAGCTGGGAGAAACATATGTCGTGACCGCCCTGCGCAAGAGACTG 348
QY 419 GAAAGGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGGATGACTCGATTGAG 478
DB 347 GAAAGGAGGAGAGACTATGTGTGAGCTTTTAAATAGAGGATGACTCGATTGAG 288
QY 479 TGATCATTAGGGCTAGGTCTGTTCTGTGGAGGTAGAGAGGCTGCTTCTGCTGGC 538
DB 287 TGATCATTAGGGCTAGGTCTGTTCTGTGGAGGTAGAGAGGCTGCTTCTGCTGGC 228
QY 539 AGGAGATGGGTTGTTGGAATCTCTAGAGGCTCTCTCGATAGGCTCGATCTG 598
DB 227 AGGAGATGGGTTGTTGGAATCTCTAGAGGCTCTCTCGATAGGCTCGATCTG 168
QY 599 GAGCAGAGCCCGAGTGTCTCTGATGATGATTTCTTCTCGAGTAGAGTTCTT 658
DB 167 GAGCAGAGCCCGAGTGTCTCTGATGATGATTTCTTCTCGAGTAGAGTTCTT 108
QY 659 TGCTTATGTTGAATTCATTCCTCTTTTCTATCAGAGATGATGTTGATTC 718
DB 107 TGCTTATGTTGAATTCATTCCTCTTTTCTATCAGAGATGATGTTGATTC 48
QY 719 TTTTGTGTCGATTAATGTTTAAAGTAAAGTTT 765
DB 47 TTTTGTGTCGATTAATGTTTAAAGTAAAGTTT 1

```

RESULT 15

```

EM977432/c
LOCUS      750 bp      mRNA      linear      EST 21-MAR-2002
DEFINITION
UI-CF-EN1-see-h-15-0-UI-s1 UI-CF-EN1 Homo sapiens cDNA clone
UI-CF-EN1-see-h-15-0-UI 3', mRNA sequence.
ACCESSION
EM977432.1 GI:19595843
VERSION
EM977432.1
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 750)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
97044477
COMMENT
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=yes.

```

FEATURES

source

Location/Qualifiers

1..750

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-CF-EN1-see-h-15-0-UI"

/clone_lib="UI-CF-EN1"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial

cells"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-CF-EN1 is a normalized cDNA library containing the

following tissue(s): Primary Lung Cystic Fibrosis

Epithelial Cells. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pRTT3-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CTGCTCAGGT.

TAG_Lib=UI-CF-EN1

TAG_Tissue=Human Lung Epithelial Cell lines untreated LPS

6hr to LPS 24h

TAG_SEQ=CTGCTCAGGT"

BASE COUNT 213 a 209 c 160 g 166 t

ORIGIN

```

Query Match      84.5%; Score 738; DB 14; Length 750;
Best Local Similarity 99.5%; Pred. No. 9.2e-125;
Matches 749; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 125 GAGAGCTGCCGGGCAATTAAGAGTACCGCTCTGAGAGAGGCTGATGCTCCGCTGC 184
DB 750 GAGAGCTGCCGGGCAATTAAGAGTACCGCTCTGAGAGAGGCTGATGCTCCGCTGC 691
QY 185 TTCCGACGAGAGAGATCTCCCTGCGCTCAAGAGTGGTGGTCTTGAGGCGG 244

```

Db 690 TTCGCCACAGAGAACTCCCTGCCCCCTGGCTGAGCTCAAGGTGTGTTCTGGCGGG 631
QY 245 CTGTTCGTATGATGTGTGATCTCTCTCTGGAGCCTGCATGTGTCTACCTGATCCGGGTG 304
Db 630 CTGTTCGTATGATGTGTGATCTCTCTCTGGAGCCTGCATGTGTCTACCTGATCCGGGTG 571
QY 305 GCACGAGAGAACACAGAGCCTGCTGCGACCGTCTGAGCTCCGAGATGACAGAGAG 364
Db 570 GCACGAGAGAACACAGAGCCTGCTGCGACCGTCTGAGCTCCGAGATGACAGAGAG 511
QY 365 CAGCTGTGAGAGACACATATGTCTGTGACCGCCCTGTGCGCAAGAGACT-GGGAGG 423
Db 510 CAGCTGTGAGAGACACATATGTCTGTGACCGCCCTGTGCGCAAGAGACTGGGAGAG 451
QY 424 GAGGGAGACTGTGTGAGCTTTTAAATAGAGGATTGACTCGGATTGTGATGATC 483
Db 450 GAGGGAGACTGTGTGAGCTTTTAAATAGAGGATTGACTCGGATTGTGATGATC 391
QY 484 ATTAGGGCTGAGGTCTGTTCTCTGGAGGATGAGACGGGTGCTTCTGTGCGAGGGA 543
Db 390 ATTAGGGCTGAGGTCTGTTCTCTGGAGGATGAGACGGGTGCTTCTGTGCGAGGGA 331
QY 544 TGGGTTTGGTTGGAATCCTTAGAGGCTCCTCTCGATGGCTGCAGTCTGGCAGC 603
Db 330 TGGGTTTGGTTGGAATCCTTAGAGGCTCCTCTCGATGGCTGCAGTCTGGCAGC 271
QY 604 AGCCCCGAGTGTTCCTCGCTGATCGATTTCTTTCCTCCAGGTAGAGTTTCTTGGCTT 663
Db 270 AGCCCCGAGTGTTCCTCGCTGATCGATTTCTTTCCTCCAGGTAGAGTTTCTTGGCTT 211
QY 664 ATGTGAATTCATGCTCCTTTCTCATCAGAAAGATGTGAAATGCTTCTTTG 723
Db 210 ATGTGAATTCATGCTCCTTTCTCATCAGAAAGATGTGAAATGCTTCTTTG 151
QY 724 TTGTCTGATTTATGTTTTTAAAGATTAACAAAGTTTTTATTAGCATTCGAAAG 783
Db 150 TTGTCTGATTTATGTTTTTAAAGATTAACAAAGTTTTTATTAGCATTCGAAAG 91
QY 784 AAGGAAAGTAAATGTACAGTTTAAATAAAAAGGGCTTCCCTTTAGAAATAATTCA 843
Db 90 AAGGAAAGTAAATGTACAGTTTAAATAAAAAGGGCTTCCCTTTAGAAATAATTCA 31
QY 844 GCATGTCTTCAAAAAAAAAAAAAAAAAAAAA 873
Db 30 GCATGTCTTCAAAAAAAAAAAAAAAAAAAAA 1

Search completed: January 10, 2003, 03:20:49
Job time : 1883.25 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:44:31 ; Search time 714.023 Seconds

(without alignments)
13246.642 Million cell updates/sec

Title: US-09-825-682A-57

Perfect score: 325

Sequence: 1 aaagggcgccgagggcct.....gtaaaaaaaaaaaaa 325

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum March 0%
Maximum March 100%
Listing first 45 summaries

Database :

```

1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_to:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vt:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rtd:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	320.4	98.6	2484	6	AX454042	AX454042 Sequence
2	320.4	98.6	2484	6	BC008765	BC008765 Homo sapi
3	307.4	94.6	2430	6	AR068043	AR068043 Sequence
4	303.4	93.4	1610	6	AX017423	AX017423 Sequence
5	303.4	93.4	53135	9	AC104742	AC104742 Homo sapi
6	303.4	93.4	148773	2	AC041029	AC041029 Homo sapi
7	289.4	89.0	2402	6	HUMSYN	105392 Human synde
8	285.8	87.9	4797	6	AX336413	AX336413 Sequence
9	285.8	87.9	4797	6	AX365741	AX365741 Sequence
10	285.8	87.9	4797	6	AX411295	AX411295 Sequence
11	285.8	87.9	4797	9	HSSYND1GN	248199 H.sapiens s
12	181.4	55.8	425	6	AX454046	AX454046 Sequence
13	144.4	44.4	492	6	AX368331	AX368331 Sequence
14	130	40.0	200548	2	AL672026	AL672026 Mus muscu
15	129.8	39.9	194985	10	AC002406	AC002406 Mouse chr
16	129.8	39.9	200548	2	AL672026	AL672026 Mus muscu
17	126.2	38.8	2410	10	S61865	S61865 syndecn-he
18	120.4	37.0	2396	10	RATSYNDECA	M81785 Rattus norv
19	120.4	37.0	176793	2	AC123473	AC123473 Rattus no
20	113	34.8	26700	6	AR068044	AR068044 Sequence
21	113	34.8	26700	6	191962	191962 Sequence 1
22	113	34.8	33934	10	MMSYNDEJA	222532 M.musculus
23	112.6	34.6	2432	6	117250	117250 Sequence 1
24	112.6	34.6	2432	10	MMSYNDE	X15487 Mouse mRNA
25	112.2	34.5	2985	10	HAMPCR	M29967 Syrian gold
26	111.4	34.3	2513	10	BC010560	BC010560 Mus muscu
27	105.8	32.6	133499	2	AC094856	AC094856 Rattus no
28	49	15.1	156533	2	AC117070	AC117070 Dictyoste
29	47.4	14.6	152409	2	PFMAL1P1	AL031744 Plasmodi
30	44.8	13.8	123280	2	AC117076	AC117076 Dictyoste
31	44.6	13.7	3565	3	ACU0481	U40481 Aplysia cal
32	43.4	13.4	14433	3	AB001369	AB001369 Plasmodi
33	43.4	13.4	121181	9	AC092821	AC092821 Homo sapi
34	43.4	13.4	180685	9	AC006432	AC006432 Homo sapi
35	43.4	13.4	240342	2	AC092862	AC092862 Homo sapi
36	43.2	13.3	180231	9	AC009161	AC009161 Homo sapi
37	43.2	13.3	318221	2	PFMAL1P3	AL049184 Plasmodi
38	42.8	13.2	87180	8	AB023044	AB023044 Arabidops
39	42.8	13.2	162200	9	AC012156	AC012156 Homo sapi
40	42.8	13.2	205952	9	AC024940	AC024940 Homo sapi
41	42.6	13.1	868	9	AK024813	AK024813 Homo sapi
42	42.6	13.1	2476	9	BC027954	BC027954 Homo sapi
43	42.6	13.1	2518	3	AF109947	AF109947 Dictyoste
44	42	12.9	1853	9	BC028120	BC028120 Homo sapi
45	41.8	12.9	3787	10	BC006627	BC006627 Mus muscu

ALIGNMENTS

```

RESULT 1
LOCUS AX454042
DEFINITION Sequence 18 from Patent WO0198539.
ACCESSION AX454042
VERSION AX454042.1 GI:21713662
KEYWORDS
SOURCE
ORGANISM Homo sapiens
            human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Mitsuhashi,M., Kambara,H., Matsumaga,H. and Kawamura,M.
TITLE Gene markers for lung cancer
JOURNAL Patent: WO 0198539-A 18 27-DEC-2001;

```


TITLE	Suppression of tumor cell growth by syndecan-1 ectodomain					
JOURNAL	Patent: US 5851993-A 3 22 DEC-1998;					
FEATURES	Location/Qualifiers 1..2430 /organism="unknown"					
BASIS COUNT	490 a	735 c	712 g	492 t	1 others	
ORIGIN						
Query Match	94.6%; Score 307.4; DB 6; Length 2430;					
Best Local Similarity	99.1%; Pred. No. 2,4e-66;					
Matches 319;	Conservative	0;	Mismatches	2;	Indels	1; Gaps 1;
Db	2100	GAGGGCGGCAAGGGGCCCTGGAGAATCCTCTGTGAACAACACCCCCTCTTGCCTGTGGGCGG	C	2158		
Qy	64	TCGCCAAGGGGCTGCTTCCTCTGGAATAATTGACAGAGGGGTTCTGTGGAGAGACTGTGCTCT		123		
Db	2159	TCTCCAGAGGGCTGCTTCTCTCTGGAATTGACAGAGGGTGCTCTTGGGACAGCTGTGCTCT		2218		
Qy	124	GAGCGCTCATCCAAAGCCAGAGTTCCTCGTTAGCTCTGTGGCCCCACCCCTGGGCCCCTG		183		
Db	2219	GAGCGCTCATCCAAAGCCAGAGTTCCTCGTTAGCTCTGTGGCCCCACCCCTGGGCCCCTG		2278		
Qy	184	GGCTGGATTCGGAATTTTTTCCAAAAGTATAGTCTTTTGCTTTTGGCAAACCTCAC		243		
Db	2279	GGCTGGATTCGGAATTTTTTCCAAAAGTATAGTCTTTTGCTTTTGGCAAACCTCAC		2338		
Qy	244	TTAATCCATAGGTTTCTCTGTCACAGTAGATTTTCCAATGTAATACTTAATATA		303		
Db	2339	TTAATCCATAGGTTTCTCTGTCACAGTAGATTTTCCAATGTAATACTTAATATA		2398		
Qy	304	AAGTAAAAAAAAAAAAAAA	325			
Db	2399	AAGTAAAAAAAAAAAAAAA	2420			
RESULT 4	AX017423 1610 bp DNA linear PAT 07-SHP-2000					
LOCUS	AX017423					
DEFINITION	Sequence 202 from Patent WO9947669.					
ACCESSION	AX017423					
VERSION	AX017423.1 GI:10042244					
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pilsbury,C.					
REFERENCE	Human nucleic acid sequences from tissue of breast tumors Patent: NO 9947669-A 202 23 SEP-1999;					
AUTHORS	SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL ENGER (DE); HINZMAN BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GBS FUER GENOMFORSCHUN (DE); PILSBURY CHRISTIAN (DE) Location/Qualifiers 1..1610					
FEATURES	source					
BASE COUNT	312 a	409 c	456 g	433 t		
ORIGIN						
Query Match	93.4%; Score 303.4; DB 6; Length 1610;					
Best Local Similarity	99.7%; Pred. No. 2,4e-65;					
Matches 304;	Conservative	0;	Mismatches	1;	Indels	0; Gaps 0;
Db	601	GAGGGCGGCAAGGGGCTGGAGAATCCTCTGTGAACAACAGCCCGCTGTGGCGGG		660		
Qy	4	GAGGGCGGCAAGGGGCTGGAGAATCCTCTGTGAACAACAGCCCGCTGTGGCGGG		63		
Db	601	GAGGGCGGCAAGGGGCTGGAGAATCCTCTGTGAACAACAGCCCGCTGTGGCGGG		660		
Qy	64	TCTCCAGAGGGCTGCTTCTCTCTGGAATTGACAGAGGGTGTCTTTGGGACAGCTGTGCTCT		123		

D6		661	TCTTAGGAGCGTCTCTTCCTCCCTGGAAATTAACAGAGGAGTGCTTTGGCACAAGCTGGACTCT	720
QY		124	GAGGCCTCATCAACAGCCAGAGTTCTCCGTTAGCTCTCTGTGGCCCCACCCTGGGCCCTTG	183
D6		721	GAGGCCTCATCAACAGCCAGAGTTCTCCGTTAGCTCTCTGTGGCCCCACCCTGGGCCCTTG	780
QY		184	GCGTGGAAATCAAGAAATTCTTGCAAGAAGAGTAGCTTTTGCTTTGGCAAAAGCTCAC	243
D6		781	GCGTGGAAATCAAGAAATTCTTGCAAGAAGAGTAGCTTTTGCTTTGGCAAAAGCTCAC	840
QY		244	TTAATCCATGGGTTTTCTCTGTACAGTAGATTTTCCAAAATGAATAACTTTAATA	303
D6		841	TTAATCCATGGGTTTTCTCTGTACAGTAGATTTTCCAAAATGAATAACTTTAATA	900
QY		304	AAGTA 308 	
D6		901	AAGTA 905 	
RESULT 5				
LOCUS		AC104792/c	53135 bp	DNA linear PRI 01-JUN-2002
DEFINITION				Homo sapiens BAC clone Rpl1-202B22 from 2, complete sequence.
ACCESSION		AC104792		
VERSION		AC104792.5	GI:20800378	
KEYWORDS		htg.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 53135) Toward a complete human genome sequence Genome Res. 8 (11), 1097-1108 (1998)		
JOURNAL		MEDLINE		
PUBMED		99063792		
REFERENCE		2 (bases 1 to 53135) Abbott,S., Kozlowski,A. and Dignan,G. The sequence of Homo sapiens BAC clone Rpl1-202B22 Unpublished (2001)		
JOURNAL		REFERENCE		
AUTHORS		3 (bases 1 to 53135) Waterston,R.H. Direct Submission Submitted (21-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
TITLE		4 (bases 1 to 53135) Waterston,R.H. Direct Submission Submitted (11-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE		5 (bases 1 to 53135) Waterston,R.H. Direct Submission Submitted (15-MAY-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
AUTHORS		6 (bases 1 to 53135) Waterston,R. Direct Submission Submitted (01-JUN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
TITLE		On May 15, 2002 this sequence version replaced gi:20136955. Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu/gsc Contact: saplen@wustl.wustl.edu Summary Statistics Center project name: H_NH0202B322		
JOURNAL		COMMENT		

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RpL1-human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Feigen, E., Tatemio, M., Cateashe, J. U. and de Jong, P. J. (1993). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.Chori.org>
 VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-698322; the clone sequenced to the right is RP11-327N17, 2000 bp overlap. Actual start of this clone is at base position 1 of RP1-202B32; actual end is at base position 98942 of RP11-327N17.

Polymorphisms exist between AC104792 and AC116171. Data from AC116171 was used to finish AC104792.

FEATURES

```

source
1..53135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-202B22"
/clone_id="RPCL11"
184..251
/rpt_family="L2"
/rpt_family="L2"
1876..1926
/rpt_family="MIR"
2632..2785
/rpt_family="WER1_type"
3411..3447
/rpt_family="(CA)n"
4149..4284
/rpt_family="WER1_type"
4410..4729
/rpt_family="WER1_type"
/rpt_family="L2"
5197..5405
/rpt_family="MIR"
5368..5573
/rpt_family="MIR"
6069..6171
/rpt_family="MIR"
7102..7181
/rpt_family="MIR"
7381..7519
/rpt_family="Alu"
8154..8517
/rpt_family="L1"
8628..8947
/rpt_family="Alu"
9450..9662
/rpt_family="L2"

```

```
repeat_region 9666. 3795 /rpt_family="L1"
repeat_region 9796. 10090 /rpt_family="Alu"
repeat_region 10091. 10207 /rpt_family="L1"
repeat_region 10383. 10521 /rpt_family="L1"
repeat_region 10624. 10907 /rpt_family="L1"
repeat_region 10970. 11376 /rpt_family="L2"
repeat_region 12019. 12609 /rpt_family="L1"
repeat_region 13339. 13387 /rpt_family="CR1"
repeat_region 13656. 13649 /rpt_family="L2"
repeat_region 19736. 19816 /rpt_family="MIR"
repeat_region 20416. 20502 /rpt_family="L2"
repeat_region 21437. 21656 /rpt_family="MIR"
repeat_region 22223. 22512 /rpt_family="Alu"
repeat_region 23163. 23280 /rpt_family="MIR"
repeat_region 23985. 24117 /rpt_family="MIR"
repeat_region 24648. 25084 /rpt_family="Alu"
repeat_region 29614. 29827 /rpt_family="MIR"
repeat_region 30410. 30602 /rpt_family="MIR"
repeat_region 32293. 32439 /rpt_family="MIR"
repeat_region 32945. 33052 /rpt_family="MIR"
repeat_region 33301. 33478 /rpt_family="CA)n"
repeat_region 33506. 33652 /rpt_family="CA)n"
repeat_region 34093. 34317 /rpt_family="L2"
repeat_region 35124. 35177 /rpt_family="CA)n"
repeat_region 35307. 35332 /rpt_family="MIR"
repeat_region 35451. 35592 /rpt_family="MIR"
repeat_region 35966. 36071 /rpt_family="CR1"
repeat_region 36872. 36894 /rpt_family="CA)n"
repeat_region 37331. 37432 /rpt_family="L2"
misc_feature 38111. 40789 /note="CGP island (%GC=71.1, o/e=0.80, #CGs=267)"
repeat_region 39015. 39091 /rpt_family="GC_rich"
repeat_region 39117. 39152 /rpt_family="GC_rich"
repeat_region 39233. 39283 /rpt_family="GC_rich"
repeat_region 41809. 41715 /rpt_family="Alu"
repeat_region 41716. 41945 /rpt_family="Alu"
repeat_region 41946. 42234 /rpt_family="L1"
repeat_region 42271. 42758 /rpt_family="L1"
```


Query Match Similarity	93.4%	Score 303.4	DB 2	Length 149773
Best Local Similarity	99.7%	Pred. No. 2.7e-65		
Matches 304	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY 4	GAGGCGGCGAGGCGCTGGAGATCTCTCTGCAAGCACGCCCTCTGCTGTGGCGCG 63			
Db 33595	GAGGCGGCGAGGCGCTGGAGATCTCTCTGCAAGCACGCCCTCTGCTGTGGCGCG 3354			
QY 64	TCTCCAGGGGCTGCTTCTCTCTGGAAATTGACGAGGGGTCTTTGGGCGAGCTGGCTCT 123			
Db 33555	TCTCCAGGGGCTGCTTCTCTCTGGAAATTGACGAGGGGTCTTTGGGCGAGCTGGCTCT 33414			
QY 124	GAGCGCTCTCCATCCACAGGCGAGGCTTCTCCGTTAGCTCCTGTGGCCCCACCTGGGCGCTG 183			
Db 33415	GAGCGCTCTCCATCCACAGGCGAGGCTTCTCCGTTAGCTCCTGTGGCCCCACCTGGGCGCTG 33474			
QY 184	GCGTGCATACAGAAATATTTTCCAAAGATGATAGCTTTTGCTTTGGCAAACTCTAC 243			
Db 33475	GCGTGCATACAGAAATATTTTCCAAAGATGATAGCTTTTGCTTTGGCAAACTCTAC 33534			
QY 244	TTATTCGAATGGGTTTTTCTCTGTGACAGTAGATTTTCCAAATGTAATTAATTAATTA 303			
Db 33535	TTATTCGAATGGGTTTTTCTCTGTGACAGTAGATTTTCCAAATGTAATTAATTAATTA 33594			
QY 304	AAGTA 308			
Db 33595	AAGTA 33599			
RESULT 7				
HUMSYN				
LOCUS	HUMSYN	2402 bp	mRNA	linear
DEFINITION	Human syndecan mRNA, complete cds.			
ACCESSION	J05392			
VERSION	J05392.1	GI:338633		
KEYWORDS	integral membrane protein; syndecan.			
SOURCE	Human breast cell line HB1-100, cDNA to mRNA, clones hsyn(4, p77).			
ORGANISM	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
REFERENCE	1 (bases 1 to 2402)			
TITLE	Malik, M., Jaakkola, P., Arvilommi, A.M. and Jalankari, M.			
JOURNAL	Sequence of human syndecan indicates a novel gene family of			
MEDLINE	integral membrane proteoglycans			
PUBMED	J. Biol. Chem. 265 (12), 6884-6889 (1990)			
COMMENT	Draft entry and printed sequence for [1] kindly submitted by			
FEATURES	M.Malik, 13-FEB-1990, for release after publication.			
SOURCE	Location/Qualifiers			
	1..2402			

QY	4	GAAGGCGGACAGGAGATCCCTCTGACAGACACGCCGCTCTGTCGACCG	63
Db	2100	GAAGGCGGACAGGAGATCCCTCTGACAGACACGCCGCTCTGTCGACCG	2158
QY	64	TCCTCAGGGGCTGCTTCTCTCTGGAAATTGACAGAGGGGTCTTTGGCAGAGATGCGCT	123
Db	2159	TCCTCAGGGGCTGCTTCTCTCTGGAAATTGACAGAGGGGTCTTTGGCAGAGATGCGCT	2218
QY	124	GAAGCGCTCCATCCAGAGCCAGGTTCTCCGTACTCTCTGACCCACCCGAGGACCTG	183
Db	2219	GAAGCGCTCCATCCAGAGCCAGGTTCTCCGTACTCTCTGACCCACCCGAGGACCTG	2278
QY	184	GGCTGGAATCAGAAATATTTCCAAAGATAGTCTTTGCTTTGGCAAACTCTAC	243
Db	2279	GGCTGGAATCAGAAATATTTCCAAAGATAGTCTTTGCTTTGGCAAACTCTAC	2338
QY	244	TTTAATCAATGGGTTTCTCTGTACAGATATTTTCCAAATGTAATAACTTTAATTA	303
Db	2339	TTTAATCAATGGGTTTCTCTGTACAGATATTTTCCAAATGTAATAACTTTAATTA	2398
QY	304	AAGT 307	
Db	2399	AAGT 2402	
RESULT 8			
AX336413			
LOCUS	AX336413	4797 bp	DNA
DEFINITION	Sequence 6922 from Patent WO0194629.	linear	PAT 09-JAN-2002
ACCESSION	AX336413		
VERSION	AX336413.1	GI:18127132	
KEYWORDS	.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,		
JOURNAL	Horrigan, S., Soppet, D.R. and Weaver, Z.		
FEATURES	Cancer gene determination and therapeutic screening using signature		
SOURCE	gene sets		
	Patent: WO 0194629-A 6922 13-DEC-2001;		
	Avalon Pharmaceuticals (US)		
	Location/Qualifiers		
	1..4797		
	Unpublished		

```

BASE COUNT      957 a 1349 c 1386 g 1099 t      6 others
ORIGIN
Query Match      87.9%; Score 285.8; DB 6; Length 4797;
Best Local Similarity 96.1%; Pred. No. 6,2e-61;
Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 GAGGGGCGAGGGGCGCTGGAATCCTCTGACAGACCGCCGCTGCGCTGGGCGCG 63
DB 3765 GAGGGGCGAGGGGCGCTGGAATCCTCTGACAGACCGCCGCTGCGCTGGGCGCG 3824
QY 64 TCTCCAGGGGCTGCTTCTCTCTGGAATTGACAGGGGCTGCTTGGGAGAGCTGCTCT 123
DB 3825 TCTCCAGGGGCTGCTTCTCTCTGGAATTGACAGGGGCTGCTTGGGAGAGCTGCTCT 3884
QY 124 GAGCGCTTCATCCAGGCGAGTTCTCGGTAGTCTGTCGCCCCACCTCGGCGCTG 183
DB 3885 GAGCGCTTCATCCAGGCGAGTTCTCGGTAGTCTGTCGCCCCACCTCGGCGCTG 3944
QY 184 GCGTGAATCAGGAATATTTTCCAAAGATGATGCTTTGGCTTTGGCAAACTCTAC 243
DB 3945 GCGTGAATCAGGAATATTTTCCAAAGATGATGCTTTGGCTTTGGCAAACTCTAC 4004
QY 244 TTAATCCAAATGGGTTTTTCTCTGTAAGATTTTCCAAATGTAATACTTAATATA 303
DB 4005 TTAATCCAAATGGGTTTTTCTCTGTAAGATTTTCCAAATGTAATACTTAATATA 4064
QY 304 AAGTA 308
DB 4065 AAGTA 4069

RESULT 9
AX365741 4797 bp DNA linear PAT 15-FEB-2002
LOCUS AX365741
DEFINITION Sequence 134 from Patent WO0200174.
ACCESSION AX365741
VERSION AX365741.1 GI:18697290
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Wang, T., Wang, A., Skeiky, Y. A., Li, S. X., Kalos, M. D., Henderson, R. A.,
McNeill, P. D., Fanger, N., Retter, M. W., Marmetakis, M., Fanger, G. R.,
Vedvyck, T. S., Carter, D., Watanabe, Y. and Peckham, D. W.
Compositions and methods for the therapy and diagnosis of lung
cancer.
Patent: WO 0200174-A 134 03-JAN-2002;
JOURNAL CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..4797
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT      957 a 1349 c 1386 g 1099 t      6 others
ORIGIN
Query Match      87.9%; Score 285.8; DB 6; Length 4797;
Best Local Similarity 96.1%; Pred. No. 6,2e-61;
Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

DB 3885 GAGCGCTTCATCCAGGCGAGTTCTCGGTAGTCTGTCGCCCCACCTCGGCGCTG 3944
QY 184 GCGTGAATCAGGAATATTTTCCAAAGATGATGCTTTGGCTTTGGCAAACTCTAC 243
DB 3945 GCGTGAATCAGGAATATTTTCCAAAGATGATGCTTTGGCTTTGGCAAACTCTAC 4004
QY 244 TTAATCCAAATGGGTTTTTCTCTGTAAGATTTTCCAAATGTAATACTTAATATA 303
DB 4005 TTAATCCAAATGGGTTTTTCTCTGTAAGATTTTCCAAATGTAATACTTAATATA 4064
QY 304 AAGTA 308
DB 4065 AAGTA 4069

RESULT 10
AX411295 4797 bp DNA linear PAT 14-JUN-2002
LOCUS AX411295
DEFINITION Sequence 3942 from Patent WO0229103.
ACCESSION AX411295
VERSION AX411295.1 GI:21444000
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Alvarez, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 3942 11-APR-2002;
JOURNAL GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
source 1..4797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/feature="EMBL/Genbank Accession No. Z48199"
/feature="EMBL/Genbank Accession No. Z48199"

BASE COUNT      957 a 1349 c 1386 g 1099 t      6 others
ORIGIN
Query Match      87.9%; Score 285.8; DB 6; Length 4797;
Best Local Similarity 96.1%; Pred. No. 6,2e-61;
Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 GAGGGGCGAGGGGCGCTGGAATCCTCTGACAGACCGCCGCTGCGCTGGGCGCG 63
DB 3765 GAGGGGCGAGGGGCGCTGGAATCCTCTGACAGACCGCCGCTGCGCTGGGCGCG 3824
QY 64 TCTCCAGGGGCTGCTTCTCTCTGGAATTGACAGGGGCTGCTTGGGAGAGCTGCTCT 123
DB 3825 TCTCCAGGGGCTGCTTCTCTCTGGAATTGACAGGGGCTGCTTGGGAGAGCTGCTCT 3884
QY 124 GAGCGCTTCATCCAGGCGAGTTCTCGGTAGTCTGTCGCCCCACCTCGGCGCTG 183
DB 3885 GAGCGCTTCATCCAGGCGAGTTCTCGGTAGTCTGTCGCCCCACCTCGGCGCTG 3944
QY 184 GCGTGAATCAGGAATATTTTCCAAAGATGATGCTTTGGCTTTGGCAAACTCTAC 243
DB 3945 GCGTGAATCAGGAATATTTTCCAAAGATGATGCTTTGGCTTTGGCAAACTCTAC 4004
QY 244 TTAATCCAAATGGGTTTTTCTCTGTAAGATTTTCCAAATGTAATACTTAATATA 303
DB 4005 TTAATCCAAATGGGTTTTTCTCTGTAAGATTTTCCAAATGTAATACTTAATATA 4064
QY 304 AAGTA 308
DB 4065 AAGTA 4069

RESULT 11
HSSYND1GN 4797 bp DNA linear PRI 09-FEB-1995
LOCUS HSSYND1GN
DEFINITION H.sapiens syndecan-1 gene (exons 2-5).
ACCESSION Z48199

```

Db	3765	GAGGGGCGCAGGGGGCTGGAGATCTCTGACAGGCTCACGGCCCTCTCTGTGGGGCGG	3824
QY	64	TTCTCAGGGGCTGTCTTCTCTCTGGAAATTGACAGGGGTGTCTTGGGACAGCTGGCTCT	123
Db	3825	TTCTCAGAGGGCTGTCTTCTCTCTGGAAATTGACAGGGGTGTCTTGGGACAGCTGGCTCT	3884
QY	124	GAGGGCTTCATCCAAAGCGCAGGTTCCTCGGTAGCTCCGTGGGCCCACTGGGCGCTG	183
Db	3885	GAGGGCTTCATCCAAAGCGCAGGTTCCTCGGTAGCTCCGTGGGCCCACTGGGCGCTG	3944
QY	184	GAGCGGATCAGGAATATTTTCCAAAGGTAGTAGTCTTTTGTCTTTTGGCAAACTCTAC	243
Db	3945	GAGCGGATCAGGAATATTTTCCAAAGGTAGTAGTCTTTTGTCTTTTGGCAAACTCTAC	4004
QY	244	TTAATCCATGGGTTTTTCTCTGTACAGTAGATTTTCCAAATGTAAATACTTTAATA	303
Db	4005	TTAATCCATGGGTTTTTCTCTGTACAGTAGATTTTCCAAATGTAAATACTTTAATA	4064
QY	304	AAAGTA 308	
Db	4065	AAAGTA 4069	
RESULT 12			
AX454046			
LOCUS	AX454046	425 bp	DNA
DEFINITION	Sequence 22 from Patent WO0198539.		linear
ACCESSION	AX454046		
VERSION	AX454046.1	GI:21713684	
KEYWORDS			
SOURCE			
ORGANISM			
synthetic construct.			
synthetic construct			
artificial sequences.			
REFERENCE			
1			
AUTHORS	Mitsunashi, M., Kambara, H., Matsunaga, H. and Kawamura, M.		
TITLE	Gene markers for lung cancer		
JOURNAL	Patent: WO 0198539-A 22-27-DEC-2001;		
	Hitachi Chemical Co., Ltd. (JP) ; HITACHI CHEMICAL RESEARCH CENTER,		
	INC. (US) ; Hitachi, Ltd. (JP)		
	Location/Qualifiers		
FEATURES			
source			
1. 425			
/organism="synthetic construct"			
/db_xref="taxon:32630"			
/note="Syndecn 1 probe."			
BASE COUNT	62 a	124 c	146 g
ORIGIN			93 t
Query Match			
Best Local Similarity	55.8%;	Score 181.4;	DB 6;
Matches 188; Conservative	0;	Pred. No. 7.5e-35;	Length 425;
		Mismatches 11;	Indels 0;
		Gaps 0;	
QY	4	GAGGGGCGCAGGGGCTGGAGATCTCTCTGACACACGCCCTCTCTGTGTGGCGCG	63
Db	227	GAGGGGCGCAGGGGCTGGAGATCTCTCTGACACACGCCCTCTCTGTGTGGCGCG	286
QY	64	TTCTCAGGGGCTGTCTTCTCTCTGGAAATTGACAGGGGTGTCTTGGGACAGCTGGCTCT	123
Db	287	TTCTCAGGGGCTGTCTTCTCTCTGGAAATTGACAGGGGTGTCTTGGGACAGCTGGCTCT	346
QY	124	GAGCGCTCATCCAAAGCGCAGGTTCCTCGGTAGCTCCGTGTGGCCCCACCTGGGCGCTG	183
Db	347	GAGCGCTCATCCAAAGCGCAGGTTCCTCGGTAGCTCTGTGTGGCCCCACCTGGGCGCTG	406
QY	184	GAGCTGGAATCAGGAATATT 202	
Db	407	GAGCTGGAATCAGGAATATT 425	
RESULT 13			
AX368331/C			
LOCUS	AX368331	492 bp	DNA
DEFINITION	Sequence 1041 from Patent WO0204514.		linear
ACCESSION	AX368331		
PAT 16-FEB-2001			

```

VERSION      AX36831.1 GI:18656404
KEYWORDS
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
AUTHORS      Wang, T., Watanabe, Y., Henderson, R.A., Johnson, J.C., Retter, M.W.,
              Manierakis, M., Carter, D., Fanger, G.R., Vedvick, T.S., Bangur, C.S.,
              Mcnabb, A., Fanger, N., Switzer, A., McNeill, P.D. and Clapper, J.D.
TITLE        Compositions and methods for the therapy and diagnosis of lung
              Cancer
JOURNAL      Patient: WO 0204514-A 1041 17-JAN-2002;
              CORNIX CORPORATION (US)
FEATURES
  source      Location/Qualifiers
              1..492
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   153 a 126 c 110 g 102 t 1 others
ORIGIN
Query Match 44.4%; Score 144.4; DB 6; Length 492;
Best Local Similarity 98.6%; Pred. No. 1.4e-25;
Matches 145; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 162 TGTGGCCGCCACCTGGGCGCTGGGATCAGGAATATTTTCCAGAGATGATGCT 221
Db 492 TGTGGCCGCCACCTGGGCGCTGGGATCAGGAATATTTTCCAGAGATGATGCT 433
Qy 222 TTGGCTTTGGCAAACTCTACTTAATCCATGAGTTTCTCTGACATAGATTTTCC 281
Db 432 TTGGCTTTGGCAAACTCTACTTAATCCATGAGTTTCTCTGACATAGATTTTCC 373
Qy 282 AAATGTAATAAATTTAATATAAGTA 308
Db 372 AAATGTAATAAATTTAATATAAGTA 346

RESULT 14
AL672026/c 200548 bp DNA linear HTG 17-AUG-2002
LOCUS      Mus musculus chromosome X clone RP23-403011, *** SEQUENCING IN
DEFINITION
ACCESSION  AL672026 200548 bp DNA linear HTG 17-AUG-2002
VERSION     AL672026.9 GI:22204360
KEYWORDS   HTG; HTGS_PHASE2; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 200548)
AUTHORS     Triomans, A.
TITLE       Direct Submission
JOURNAL     Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk; Clone Requests: clonerequests@sanger.ac.uk
            On Aug 11, 2002 this sequence version replaced gi:21955546.
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: BM403011
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 5% of reads
            Consensus quality: 200513 bases at least Q40
            Consensus quality: 200548 bases at least Q30
            Consensus quality: 200548 bases at least Q20
            Insert size: 200548; sum-of-contigs
            Insert size: 172347; 7.0% error; agarose-fp
COMMENT

```

```

FEATURES
  source      Location/Qualifiers
              1..200548
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /chromosome="X"
              /clone="RP23-403011"
              /clone_lib="RPC1-23"
              misc_feature 1..200548
              /note="assembly_fragment:03152"
BASE COUNT   56146 a 44414 c 44985 g 55003 t
ORIGIN
Query Match 40.0%; Score 130; DB 2; Length 200548;
Best Local Similarity 76.1%; Pred. No. 6.4e-22;
Matches 216; Conservative 0; Mismatches 55; Indels 13; Gaps 4;
Qy 38 CCAAGCCCTCTGCTGCTGAGCGGCTCTCCAGAGGCTCTTCTCTCGGAATTGACGA 97
Db 17768 CCAAGGCTCTCCACCTTTGGTACCATCTCTAGTCAC-CTTTCTCTCGGAATTGACGA 177710
Qy 98 GGGGTCTCTGGGCAAGAGCTGCTGAGCGCTTCATCCAGGCGAGTTCCTCGT 157
Db 177709 GACACATCTTAAGTATGCTGAGCAGCTG-CTTCTCCATCAAGAACCAAGTTCATCTTCAG 177651
Qy 158 CTCCTGTGCCCCCAGCCCGGCGCTGGGATCAGGAATATTTTCCAAAGATGATA 217
Db 177650 CTCCTGTGCCCCCAG-CCCCAGGCTGAGTCAAAATTTTCCCAAAGATGATA 177597
Qy 218 GCTTTTGTCTTTGGCAAACTCTACTTAATCCAAATGGGTTTCTCTGACATAGATT 277
Db 177596 GCTTTTGTCTTTGGCAAAAGCTACTTAATCCAAATGGGTTTCTCTGACATAGATT 177542
Qy 278 TTCCAAATGTAATAAATTTAATATAAGTAAAAA 321
Db 177541 TTCCAAATGTAATAAATTTAATATAAGTAAAAA 177498

RESULT 15
AC002406 194985 bp DNA linear ROD 21-MAR-1998
LOCUS      Mouse chromosome X BAC B178A13 (Research Genetics mouse BAC
DEFINITION
ACCESSION  AC002406 194985 bp DNA linear ROD 21-MAR-1998
VERSION     AC002406 complete sequence.
KEYWORDS
SOURCE      HTG.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 194985)
AUTHORS     Muzny, D., Aronson, A.D., Brundage, E., Carvelli, K., Chen, E., Chen, J.,
            Di, W., Ding, Y., Dugan, S., Durbin, J., Forcum, J., Ganes, R.,
            Garcia, C., Goodman, M., Gorrell, J.H., Haywood, M., Jackson, L.,
            Jin, S., Kampal, R., Karpach, S., Leal, B., Li, Y., Liu, W., Logan, O.,
            Lu, J., Ly, T., Martinez, C., Oswal, G., Perez, L., Rashid, N.D.,
            Rowland, K., Savage, L., Scherer, S.S., Shen, H., Timms, K.M., Todd, J.,
            Vo, Q., Worley, K.C., Yu, W., Chnault, C., Nelson, D. and Gibbs, R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 194985)
AUTHORS     Chin, M.W.
TITLE       Direct Submission
JOURNAL     Submitted (11-AUG-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 194985)
AUTHORS     Worley, K.C.

```

TITLE
JOURNAL

Direct Submission
Submitted (21-MAR-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Mar 23, 1998 this sequence version replaced gi:2909685.

Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES

source
1. 194985
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="B178A13"
/clone_lib="Research Genetics mouse BAC library"
226.352
repeat_region
562.599
/rpt_family="MSTD"
repeat_region
/rpt_family="(TA)n"
complement(863.1868)
repeat_region
/rpt_family="L1_MM"
complement(2661.3921)
repeat_region
/rpt_family="B4A"
complement(5160.5372)
repeat_region
/rpt_family="B3"
complement(3456.5705)
repeat_region
/rpt_family="L1MD3"
complement(5766.5807)
repeat_region
/rpt_family="(TAAA)n"
complement(5900.6088)
repeat_region
/rpt_family="L1MD3"
complement(6433.6546)
repeat_region
/rpt_family="(TAAA)n"
complement(6547.6575)
repeat_region
/rpt_family="AT-rich"
complement(6576.6776)
repeat_region
/rpt_family="L1MD2"
complement(6808.6947)
repeat_region
/rpt_family="B1-F"
complement(7019.7348)
repeat_region
/rpt_family="L1MD2"
7396.7541
repeat_region
/rpt_family="B1_MM"
7542.7587
repeat_region
/rpt_family="(CAA)n"
complement(7915.8149)
repeat_region
/rpt_family="L1M4"
8166.8512
repeat_region
/rpt_family="L1M4"
complement(8730.9162)
repeat_region
/rpt_family="L1_MM"
9149.9265
repeat_region
/rpt_family="L1_MM"
complement(9454.9606)
repeat_region
/rpt_family="RLTR10"
complement(9734.10166)
repeat_region
/rpt_family="L1_MM"
11635.11661
repeat_region
/rpt_family="POLY_A"
complement(11830.11966)
repeat_region
/rpt_family="RLTR10"

repeat_region 11967.12203
/rpt_family="L1_MM"
repeat_region 12575.12708
/rpt_family="B1-F"
repeat_region 12711.12772
/rpt_family="(CAAAA)n"
complement(13570.13598)
repeat_region 14287.14359
/rpt_family="AT-rich"
/rpt_family="MIR"
repeat_region 14418.15428
/rpt_family="L1"
repeat_region 15490.15596
/rpt_family="L1"
repeat_region 15828.16199
/rpt_family="LX4"
complement(16753.16830)
repeat_region 17198.17275
/rpt_family="MTE"
repeat_region 17311.17443
/rpt_family="MLT1A1"
complement(19559.19879)
repeat_region 19964.26215
/rpt_family="MTE"
/rpt_family="L1F"
complement(26283.26511)
repeat_region 26468.27422
/rpt_family="L1_MM"
complement(27624.27755)
repeat_region 27759.28051
/rpt_family="B1-F"
repeat_region 28280.28490
/rpt_family="MTC"
/rpt_family="L1"
repeat_region 28491.33172
/rpt_family="L1_MM"
33178.33672
/note="Region: Similar to U29402|NM029402 Mus musculus acidic ribosomal phosphoprotein P1 mRNA, complete cds"
repeat_region 33682.35079
/rpt_family="L1_MM"
repeat_region 37207.37292
/rpt_family="RSINE1"
complement(37310.37338)
repeat_region 37354.37474
/rpt_family="(CAT)n"
complement(37574.37728)
repeat_region 37574.37728
/rpt_family="RSINE1"
complement(38061.38179)
repeat_region 39320.39492
/rpt_family="(TA)n"
/rpt_family="B4"
repeat_region 39603.39661
/rpt_family="(CA)n"
complement(40026.40452)
repeat_region 40580.40624
/rpt_family="LX2B"
/rpt_family="(CA)n"
complement(40951.41040)
repeat_region 41143.41538
/rpt_family="MLT2B"
/rpt_family="MTE"
complement(42185.42262)
repeat_region 43208.43318
/rpt_family="(CAG)n"
complement(43486.43974)
repeat_region 43486.43974
/rpt_family="L1"

```
repeat_region      43954. .44894  
/rpt_family="L1_MM"  
repeat_region      45112. .45198  
/rpt_family="(CAA)n"  
repeat_region      complement(45459. .45479)  
/rpt_family="AT_rich"  
repeat_region      46721. .46836  
/rpt_family="(GAA)n"  
repeat_region      46914. .47150  
/rpt_family="MTD"  
repeat_region      47213. .47296  
/rpt_family="PB1D10"  
repeat_region      50241. .50324  
/rpt_family="MTC"  
repeat_region      50328. .51555  
/rpt_family="L1_MM"  
repeat_region      51568. .52378  
/rpt_family="L1_MM"  
repeat_region      52399. .52434  
/rpt_family="POLY_A"  
repeat_region      52446. .52503  
/rpt_family="L1_MM"  
repeat_region      52520. .52838  
/rpt_family="MTC"  
repeat_region      53200. .53245  
/rpt_family="(AAAA)n"  
repeat_region      complement(53474. .53534)
```

Query Match 39.9%; Score 129.8; DB 10; Length 194985;
Best Local Similarity 75.6%; Pred. No. 7.1e-22;
Matches 217; Conservative 0; Mismatches 57; Indels 13; Gaps 4;

```
QY 38 CCACGCCCGTCTGCGCTGTGGGCGCGCTCCAGGGGCTGCTCTCTCTGGAATTAACGA 97  
Db 148547 CCAAGCCTCTCCACCTTGACATCTCTAGTAC -CCTTCTCCCGAAGTTGACAA 148605  
QY 98 GGGGTGCTGGGCGAAGTGGCTCTGAGCGCCTCCATCCAGGCCAGTTCTCCGTTAG 157  
Db 148606 GACACATCTTGAGTATGGCTGGCACTG -GTTCCTCATCAGAACCAAGTTCACTTCAG 148664  
QY 158 CTCTGTGGGCCCACTGGGCGCTGGGCTGGAATCAGGATATTTTCCAAAGAGTATA 217  
Db 148665 CTCTGTGGGCCCGCCT-----CGAGCTGGAGTCAGAAATGTTCCCAAGAGTGAGA 148718  
QY 218 GTCTTTTGCTTTTGGCAAAACTTACTTAATCAATGGGTTTCTCTGTACAGTATGATT 277  
Db 148719 GTCTTTGCTTTTGGCAAAACGCTACTTAATCAATGGGT-----TATGTACGTAGATT 148773  
QY 278 TTCAAATGTAATTAATTTAATATAAGTAAAAAAGAGAGAGAGAGAGAGAGAGAGAG 324  
Db 148774 TTGCAGATGTAATTAATTTAATATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 148820
```

Search completed: January 10, 2003, 02:45:16
Job time : 877.023 secs

XX 13-JUN-1995; 95NO-FI00344.
 PF 13-JUN-1994; 94US-0258862.
 PR 13-JUN-1994; 94US-0258862.
 XX
 XX (JALK/) JALKANEN M.
 PA (MALI/) MALI M.
 XX
 XX Jalkanen M, Mali M;
 PI
 XX WPI: 1996-049416/05.
 DR P-PSDB; AAR87000.
 XX
 XX New syndecan ectodomain compsn. to reduce tumour growth - for
 PT delivery to extracellular environment for suppressing tumour growth
 PT in malignant and non-malignant tumours
 XX
 XX Disclosure: Fig 1; 49pp; English.
 PS
 XX A human DNA (AAT08125) codes for syndecan-1 (AAR87000), a cell surface
 CC proteoglycan. The ectodomain of syndecan-1 (amino acids 18-251)
 CC is capable of restoring a more differentiated morphology to tumour
 CC cells and to suppress the growth of malignant cells. The DNA
 CC can be used for the prodn. of recombinant ectodomain, pref. using
 CC host cells of a similar cell type as that of the tumour. The
 CC mouse syndecan-1 gene (AAT08126) has also been identified.
 XX
 SQ Sequence 2430 BP; 490 A; 736 C; 711 G; 492 T; 1 other;
 Query Match 94.6%; Score 307.4; DB 17; Length 2430;
 Best Local Similarity 99.1%; Pred. No. 4.5e-68;
 Matches 319; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 QY 4 GAGGGGGGAGGGGGCTGGAATCTCTGACAGACCCGCTGCTGAGGCGG 63
 Db 2100 GAGGGGGGAGGGGGCTGGAATCTCTGACAGACCCGCTGCTGAGGCGG 2158
 QY 64 TCTCCAGGGGGCTGCTCTCTGGAATTGACAGGGGGTGTCTTGAGCAGAGCTGCTCT 123
 Db 2159 TCTCCAGGGGGCTGCTCTCTGGAATTGACAGGGGGTGTCTTGAGCAGAGCTGCTCT 2218
 QY 124 GAGCGCTTCATCCAGGCGCAGGTTCTCGTTAGCTCTGAGGCCCACTTGGGCGCTG 183
 Db 2219 GAGCGCTTCATCCAGGCGCAGGTTCTCGTTAGCTCTGAGGCCCACTTGGGCGCTG 2278
 QY 184 GGCCTGAATCAGGAATATTTTCCAAAGAGTAGTCTTTGGCAAACTCTAC 243
 Db 2279 GGCCTGAATCAGGAATATTTTCCAAAGAGTAGTCTTTGGCAAACTCTAC 2338
 QY 244 TTAATCCAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 303
 Db 2339 TTAATCCAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 2398
 QY 304 AAGTAAAAAAAAAAAAAAAA 325
 Db 2399 AAGTAAAAAAAAAAAAAAAA 2420
 RESULT 2
 ID AAV81282 standard; DNA; 2430 BP.
 XX AAV81282;
 AC
 XX 11-MAR-1999 (first entry)
 DT
 XX Human syndecan-1 encoding DNA.
 DB
 XX S ndecan-1; tumour; ectodomain; epithelial; mesenchymal; breast; human;
 KM endometrial tumour; prostatic carcinoma; oestrogenic; androgenic; steroid;
 KM glioma; myeloma; carcinoma; sarcoma; lymphoma; adenoma; ds.
 XX
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH 206..1138
 FT CDS
 FT
 FT polyA_signal
 FT 2383..2388
 FT /*tag= b
 FT /*note= "q"
 PN US585193-A.
 XX
 XX 22-DEC-1998.
 PD
 XX
 PF 07-JUN-1995; 95US-0488139.
 XX
 XX 07-JUN-1995; 95US-0488139.
 PR 13-JUN-1994; 94US-0258862.
 XX
 XX (BIOT-) BIOTIE THERAPIES LTD.
 PA
 XX Jalkanen M, Mali M;
 PI
 XX WPI: 1999-104635/09.
 DR P-PSDB; AAW95197.
 XX
 PT Reducing growth of tumour cells - with ectodomain of syndecan
 PT applied to the extracellular environment to induce a more
 PT differentiated phenotype, particularly for hormone-dependent breast,
 PT endometrial or prostatic cancers
 XX
 PS Claim 7; Fig 1A-B; 48pp; English.
 XX
 CC This DNA encodes a human syndecan-1 protein. This is used in the method
 CC of the invention of reducing the growth of tumour cells. The method
 CC comprises supplying a syndecan ectodomain to the environment around the
 CC cells causing them to develop a more differentiated phenotype. The method
 CC is used to suppress tumour cells of epithelial, mesenchymal, pre-B or
 CC plasma cell origin, especially breast, endometrial or prostatic tumours,
 CC and particularly those responsive to an oestrogenic or androgenic
 CC steroid. More generally it can be used to treat malignant or non-
 CC malignant tumours, particularly those characterised by loss of syndecan,
 CC e.g. gliomas, myelomas, carcinomas, sarcomas, lymphomas and adenomas.
 XX
 SQ Sequence 2430 BP; 490 A; 735 C; 712 G; 492 T; 1 other;
 Query Match 94.6%; Score 307.4; DB 20; Length 2430;
 Best Local Similarity 99.1%; Pred. No. 4.5e-68;
 Matches 319; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 QY 4 GAGGGGGGAGGGGGCTGGAATCTCTGACAGACCCGCTGCTGAGGCGG 63
 Db 2100 GAGGGGGGAGGGGGCTGGAATCTCTGACAGACCCGCTGCTGAGGCGG 2158
 QY 64 TCTCCAGGGGGCTGCTCTCTGGAATTGACAGGGGGTGTCTTGAGCAGAGCTGCTCT 123
 Db 2159 TCTCCAGGGGGCTGCTCTCTGGAATTGACAGGGGGTGTCTTGAGCAGAGCTGCTCT 2218
 QY 124 GAGCGCTTCATCCAGGCGCAGGTTCTCGTTAGCTCTGAGGCCCACTTGGGCGCTG 183
 Db 2219 GAGCGCTTCATCCAGGCGCAGGTTCTCGTTAGCTCTGAGGCCCACTTGGGCGCTG 2278
 QY 184 GGCCTGAATCAGGAATATTTTCCAAAGAGTAGTCTTTGGCAAACTCTAC 243
 Db 2279 GGCCTGAATCAGGAATATTTTCCAAAGAGTAGTCTTTGGCAAACTCTAC 2338
 QY 244 TTAATCCAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 303
 Db 2339 TTAATCCAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 2398
 QY 304 AAGTAAAAAAAAAAAAAAAA 325
 Db 2399 AAGTAAAAAAAAAAAAAAAA 2420

```

RESULT 3
AAFI7989
ID AAF17989 standard, DNA; 1763 BP.
XX
AC AAF17989;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 8.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnery;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN W0200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
XX
P-PSDB; AAB58113.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 1; Page 495; 1425bp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1763 BP; 344 A; 450 C; 499 G; 470 T; 0 other;
XX
Query Match 89.7%; Score 291.4; DB 21; Length 1763;
Best Local Similarity 99.3%; Pred. No. 4,5e-64;
Matches 303; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
XX
QY 4 GAGGGCGGCAAGGGGCTGAGATCTCTGACAGACACGCCCGCTCTGCGCCG 63
DB 752 GAGGGCGGCAAGGGGCTGAGATCTCTGACAGACACGCCCGCTCTGCGCCG 811
QY 64 TCTCCAGAGGCTCTCTCTCCGGAATTTGACGAGGGGTCTCTTGCGGAGAGCTGGTCT 123
DB 812 TCTCCAGAGGCTCTCTCTCCGGAATTTGACGAGGGGTCTCTTGCGGAGAGCTGGTCT 871

```

```

QY 124 GAGGCGCTCCATCCAGGCGCAGGTTCTCCGTTAGCTCCTGTGCGCCGACCTGGGCGCTTG 183
DB 872 GAGGCGCTCCATCCAGGCGCAGGTTCTCCGTTAGCTCCTGT -GCCCGACCTGGGCGCTTG 930
QY 184 GCGTGAATCAGGAATATTTTCCAAAGATGATAGTCTTTTGGCAAAACTCTAC 243
DB 931 GCGTGAATCAGGAATATTTTCCAAAGATGATAGTCTTTTGGCAAAACTCTAC 990
QY 244 TTAATCCATGAGTTTCTCTGTACAGTAGATTTTCCAAATGTAATTAATTAATA 303
DB 991 TTAATCCATGAGTTTCTCTGTACAGTAGATTTTCCAAATGTAATTAATTAATA 1050
QY 304 AAGTA 308
DB 1051 AAGTA 1055

RESULT 4
AAS01308
ID AAS01308 standard, cDNA; 320 BP.
XX
AC AAS01308;
XX
DT 16-JUL-2001 (first entry)
XX
DE Bladder cancer-associated sequence, TCC75E3.
XX
KW Bladder; cancer; transitional cell carcinoma; diagnostic; TCC75E3; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH FT polyA_site 311
FH FT /*tag= a
XX
PN W0200122864-A2.
XX
PD 05-APR-2001.
XX
PF 27-SEP-2000; 2000WO-US41005.
XX
PR 27-SEP-1999; 99US-0156153.
XX
PA (QUR-) QURAK BIOTECH INC.
PA (KOHN/) KOHN K I.
XX
XX Feinstein E, Mor O;
XX
DR WPI; 2001-258076/26.
XX
PT New polynucleotide sequences upregulated in bladder cancer for
PT diagnosing bladder cancer and inhibition of expression is useful for
PT treating and regulating bladder cancer-associated pathologies -
XX
PS Claim 4; Page 56; 64pp; English.
XX
CC The sequence represents the coding sequence of bladder cancer-associated
CC sequence, transitional cell carcinoma clone, TCC75E3. The sequence is
CC upregulated in bladder cancer and its expression is indicative of bladder
CC cancer. The sequence can be used as a marker, and can be used for
CC diagnosing bladder cancer. Antibodies, ribozymes, antisense
CC oligonucleotides or a dominant negative peptide directed against the
CC sequence are useful for regulating bladder cancer-associated pathologies
CC in a patient.
XX
SQ Sequence 320 BP; 73 A; 82 C; 81 G; 84 T; 0 other;
XX
Query Match 88.4%; Score 287.2; DB 22; Length 320;
Best Local Similarity 97.8%; Pred. No. 2.9e-63;
Matches 313; Conservative 0; Mismatches 3; Indels 4; Gaps 2;
QY 1 AAAGAGGCGCGAGGGGCTGAGATCTCTGACAGACACGCCCGTCTCTGCGCG 60

```


CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_poc_sequences.
 XX
 XX Sequence 4797 BP; 957 A; 1349 C; 1386 G; 1099 T; 6 other;
 SO
 Query Match 87.3%; Score 285.8; DB 24; Length 4797;
 Best Local Similarity 96.1%; Pred. No. 1.6e-62;
 Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 4 GAGGGGGGAGGGGGGCTGGAGATCCCTGACAGACCGCCGCTGCTGAGGCGG 63
 Db 3765 GAGGGGGGAGGGGGCTGGAGATCCCTGACAGGCTCAGCCGCTGCTGAGGCGG 3824
 QY 64 TCTCCAGGGGCTGCTTCTCTCTGAAATTGACAGGGGCTGCTGAGAGCTGGCTCT 123
 Db 3825 TCTCCAGGGGCTGCTTCTCTCTGAAATTGACAGGGGCTGCTGAGAGCTGGCTCT 3884
 QY 124 GACGGCTTCATCCAGGCCAGGCTTCCTGTTAGCTCTGAGGCCACCCCTGGGCGCTG 183
 Db 3885 GACGGCTTCATCCAGGCCAGGCTTCCTGTTAGCTCTGAGGCCACCCCTGGGCGCTG 3944
 QY 184 GGTGGAAATCAGAAATATTTTCCAAAGAGTAGTCTTTGGCTTTGGCAAACTTAC 243
 Db 3945 GGTGGAAATCAGAAATATTTTCCAAAGAGTAGTCTTTGGCTTTGGCAAACTTAC 4004
 QY 244 TTATTCATAGGCTTTTCTCTGTCAGTAGATTTTCCAAATGTATTAATCTTATATA 303
 Db 4005 TTATTCATAGGCTTTTCTCTGTCAGTAGATTTTCCAAATGTATTAATCTTATATA 4064
 QY 304 AAGTA 308
 Db 4065 AAGTA 4069
 AC ABL6585;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Kidney cancer related gene sequence SEQ ID NO:6922.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytoskeletal; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-231133P.
 PR 18-SEP-2000; 2000US-231617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.

PR 20-SHP-2000; 2000US-234052P.
 PR 22-SHP-2000; 2000US-234509P.
 PR 22-SHP-2000; 2000US-234567P.
 PR 25-SHP-2000; 2000US-234923P.
 PR 25-SHP-2000; 2000US-234924P.
 PR 25-SHP-2000; 2000US-235077P.
 PR 25-SHP-2000; 2000US-235082P.
 PR 25-SHP-2000; 2000US-235134P.
 PR 25-SHP-2000; 2000US-235280P.
 PR 26-SHP-2000; 2000US-235637P.
 PR 26-SHP-2000; 2000US-235638P.
 PR 27-SHP-2000; 2000US-235711P.
 PR 27-SHP-2000; 2000US-235720P.
 PR 27-SHP-2000; 2000US-235840P.
 PR 27-SHP-2000; 2000US-235863P.
 PR 28-SHP-2000; 2000US-236028P.
 PR 28-SHP-2000; 2000US-236032P.
 PR 28-SHP-2000; 2000US-236033P.
 PR 28-SHP-2000; 2000US-236034P.
 PR 28-SHP-2000; 2000US-236109P.
 PR 28-SHP-2000; 2000US-236111P.
 PR 28-SHP-2000; 2000US-236842P.
 PR 29-SHP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237356P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endreess G, Horrihan S;
 PI Sopet DR, Weaver Z;
 XX
 DR WP1; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 XX
 ES Claim 1; SEQ ID 6922; 44pp; English.
 XX
 XX The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 847 sequences (given in ABL6164
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytoskeletal
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.
 XX
 SO Sequence 4797 BP; 957 A; 1349 C; 1386 G; 1099 T; 6 other;
 Query Match 87.3%; Score 285.8; DB 24; Length 4797;
 Best Local Similarity 96.1%; Pred. No. 1.6e-62;
 Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 GAGGGGCGCAGGGGCGCTGGAGATCTCTCTGACAGACCGCCGCTCTGCTGTGGGCGCG 63
 Db 3765 GAGGGGCGCAGGGGCGCTGGAGATCTCTCTGACAGACCGCCGCTCTGCTGTGGGCGCG 3824
 QY 64 TCTCAGAGGGGCTGCTCTCTCTGGAATTTGACAGAGGGGTGTCTTGGGACAGCTGGCTCT 123
 Db 3825 TCTCAGAGGGGCTGCTCTCTCTGGAATTTGACAGAGGGGTGTCTTGGGACAGCTGGCTCT 3884
 QY 124 GAGGGGCTCTCATCCAGGCGCAGGTTCTCGCTTACCTCTGTGGGCCCGACCCCTGGGCGCTG 183
 Db 3885 GAGGGGCTCTCATCCAGGCGCAGGTTCTCGCTTACCTCTGTGGGCCCGACCCCTGGGCGCTG 3944
 QY 184 GGCTGGAAATCAGGAATATTTTCCAAAGAGTGAATCTTTTGGGCAAACTCTAC 243
 Db 3945 GGCTGGAAATCAGGAATATTTTCCAAAGAGTGAATCTTTTGGGCAAACTCTAC 4004
 QY 244 TTAATCCAAATGGGTTTTTCTCTGTACAGTATTTTCCAAATGTAATACTTTAATA 303
 Db 4005 TTAATCCAAATGGGTTTTTCTCTGTACAGTATTTTCCAAATGTAATACTTTAATA 4064
 QY 304 AAGTA 308
 Db 4065 AAGTA 4069

RESULT 8

ABL49089
 ID ABL49089 standard; cDNA; 4797 BP.

AC ABL49089;

DT 01-MAY-2002 (first entry)

DE Human lung tumour cDNA sequence for contig 31 SEQ ID NO:134.

KX Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;

KW immune response; ss.

OS Homo sapiens.

PN WO200200174-A2.

PD 03-JAN-2002.

PF 28-JUN-2001; 2001WO-US21065.

PR 28-JUN-2000; 2000US-0606421.

PR 02-AUG-2000; 2000US-0630940.

PR 21-AUG-2000; 2000US-0643597.

PR 15-SEP-2000; 2000US-0662786.

PR 09-OCT-2000; 2000US-0685696.

PR 12-DEC-2000; 2000US-0735705.

PR 07-MAY-2001; 2001US-0850716.

PA (CORI-) CORIXA CORP.

PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;

PI McNeill PD, Fanger N, Retter MW, Narnierakis M, Fanger GR;

PI Vedvick TS, Carter D, Matanabe Y, Pecham DW;

DR WPI; 2002-090513/12.

PS Example 3; Page 238-240; 374pp; English.

CC express the lung tumour proteins are useful for treating lung cancer or
 CC stimulating an immune response. ABL49859 to ABL49300 and ABL74946 to
 CC ABL75070 represent sequences used in the exemplification of the present
 CC invention.

XX Sequence 4797 BP; 957 A; 1349 C; 1386 G; 1099 T; 6 other;
 Query Match 87.9%; Score 285.8; DB 24; Length 4797;
 Best Local Similarity 96.1%; Pred. No. 1,6e-62;
 Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 GAGGGGCGCAGGGGCGCTGGAGATCTCTCTGACAGACCGCCGCTCTGCTGTGGGCGCG 63
 Db 3765 GAGGGGCGCAGGGGCGCTGGAGATCTCTCTGACAGACCGCCGCTCTGCTGTGGGCGCG 3824
 QY 64 TCTCAGAGGGGCTGCTCTCTCTGGAATTTGACAGAGGGGTGTCTTGGGCAAACTCTAC 123
 Db 3825 TCTCAGAGGGGCTGCTCTCTCTGGAATTTGACAGAGGGGTGTCTTGGGCAAACTCTAC 3884
 QY 124 GAGGGGCTCTCATCCAGGCGCAGGTTCTCGCTTACCTCTGTGGGCCCGACCCCTGGGCGCTG 183
 Db 3885 GAGGGGCTCTCATCCAGGCGCAGGTTCTCGCTTACCTCTGTGGGCCCGACCCCTGGGCGCTG 3944
 QY 184 GGCTGGAAATCAGGAATATTTTCCAAAGAGTGAATCTTTTGGGCAAACTCTAC 243
 Db 3945 GGCTGGAAATCAGGAATATTTTCCAAAGAGTGAATCTTTTGGGCAAACTCTAC 4004
 QY 244 TTAATCCAAATGGGTTTTTCTCTGTACAGTATTTTCCAAATGTAATACTTTAATA 303
 Db 4005 TTAATCCAAATGGGTTTTTCTCTGTACAGTATTTTCCAAATGTAATACTTTAATA 4064
 QY 304 AAGTA 308
 Db 4065 AAGTA 4069

RESULT 9

AAT23809
 ID AAT23809 standard; cDNA to mRNA; 287 BP.

AC AAT23809;

DT 15-AUG-1996 (first entry)

DE Human gene signature HDWG505718.

KM Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KM human; cloning; mapping; non-biased library; diagnosis; detection;

OS Homo sapiens.

PN WO9514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; 94WO-JP01916.

PR 12-NOV-1993; 93JP-0355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI; 1995-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PS Claim 1; Page 1460; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAI19901-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX

Sequence 287 BP; 60 A; 78 C; 67 G; 80 T; 2 other;

Query Match 86.2%; Score 280.2; DB 16; Length 287;
 Best Local Similarity 98.3%; Pred. No. 1.7e-61;
 Matches 282; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 24 GATCCTCTCAGACACCGCCGCTGCTGTGACGCTCCAGGAGGCTCTCTC 83
 Db 1 GATCCTCTCAGACACCGCCGCTGCTGTGACGCTCCAGGAGGCTCTCTC 60
 QY 84 CTGAAATTCACAGGGGTGTCTTGCGACAGCTGAGCTCTGAGCGCTCCATCCAGGCC 143
 Db 61 CTGAAATTCACAGGGGTGTCTTGCGACAGCTGAGCTCTGAGCGCTCCATCCAGGCC 120
 QY 144 AGCTTCCGCTAGCTCTCTGCTGCGCCCACTGAGCGCTGAGCTGGAATCAGGAATTT 203
 Db 121 AGCTTCCGCTAGCTCTCTGCTGCGCCCACTGAGCGCTGAGCTGGAATCAGGAATTT 180
 QY 204 TCCAAAGAGTAGTCTTTTCTGCTTTGGCAAACTCTACTTAATCCAAATGAGTTTCT 263
 Db 121 TCCAAAGAGTAGTCTTTTCTGCTTTGGCAAACTCTACTTAATCCAAATGAGTTTCT 240
 QY 264 CTGTACAGTAGATTTTCCAAATGTATTAACCTTTAATATTAAGTAAA 310
 Db 241 CTGTACAGTAGATTTTCCAAATGTATTAACCTTTAATATTAAGTAAA 287

RESULT 10
 AA224631
 ID AA224631 standard; cDNA; 4787 BP.

XX AC AA224631;
 XX DT 07-DEC-1999 (first entry)
 XX DE Human lung tumor associated polynucleotide.
 XX KN Human; lung tumor; lung cancer; T cell stimulation; ss.
 XX OS Homo sapiens.
 XX PN WO9947674-A2.
 XX PD 23-SEP-1999.
 XX PF 17-MAR-1999; 99WO-US05798.
 XX PR 18-MAR-1998; 98US-0040802.
 XX PR 18-MAR-1998; 98US-0040984.
 XX PR 27-JUL-1998; 98US-0123912.
 XX PR 27-JUL-1998; 98US-0123933.
 XX PA (CORI-) CORIXA CORP.
 XX PI Read SG, Wang T;
 XX

DR WPI; 1999-571839/48.

XX New isolated lung tumor polynucleotides, used to develop products for
 PT the treatment, prevention and monitoring the progression of lung cancer
 PT
 XX Claim 12; Page 122-123; 148pp; English.

CC The invention provides isolated human lung tumor nucleic acids and
 CC polypeptides. The polypeptides can be used for the treatment of lung
 CC cancer. The polypeptides and polynucleotides can be used to stimulate T
 CC cells or antigen presenting cells for use in the treatment of lung
 CC cancer. The polypeptides and monoclonal antibodies specific for the
 CC polypeptides can also be used to inhibit the development of lung cancer.
 CC Agents which bind the polypeptides can be used for detecting lung cancer
 CC and for monitoring the progression of lung cancer.
 XX

Sequence 4787 BP; 957 A; 1346 C; 1386 G; 1092 T; 6 other;

Query Match 83.0%; Score 269.8; DB 20; Length 4787;
 Best Local Similarity 95.1%; Pred. No. 1.8e-58;
 Matches 290; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

QY 4 GAGGCGGCGAGGGGCGCTGAGATCCTCTGACAGCACGCGCTCTGCTGAGCGCG 63
 Db 3764 GAGGCGGCGAGGGGCGCTGAGATCCTCTGAGAGCTCAGCGCCGCTCTGCTGAGCGCG 3823
 QY 64 TCTCCAGGGCTCTCTCTCTGGAATTCACAGGGGTGTCTTGCGACAGCTGAGCTCT 123
 Db 3824 TCTCCAGGGCTCTCTCTCTGGAATTCACAGGGGTGTCTTGCGACAGCTGAGCTCT 3883
 QY 124 GAGGCGCTCCATCCAGAGGCGAGCTTCCCTGTAGCTCTGTGAGCCCACTGAGCGCTG 183
 Db 3884 GAGGCGCTCCATCCAGAGGCGAGCTTCCCTGTAGCTCTGTGAGCCCACTGAGCGCTG 3943
 QY 184 GAGTGAATGAGATATTTTCCAAAGAGTAGTACTTTTCTTTGGCAAACTCTAC 243
 Db 3944 GAGTGAATGAGATATTTTCCAAAGAGTAGTACTTTTCTTTGGCAAACTCTAC 4003
 QY 244 TTAATCCAAATGAGTTTCTCTGTACAGTAGATTTTCCAAATGTATTAACCTTTAATATA 303
 Db 4004 TTAATCCAAATGAGTTTCTCTGTACAGTAGATTTTCCAAATGTATTAACCTTTAATATA 4060
 QY 304 AAGTA 308
 Db 4061 AAGTA 4065

RESULT 11

AA172270
 ID AA172270 standard; DNA; 425 BP.

XX AC AA172270;
 XX DT 15-APR-2002 (first entry)
 XX DE Exons 2-6 of Syndecan 1 gene.
 XX KN Lung; cancer; metastasis; solid tumour; blood; bone marrow; syndecan 1;
 XX collagen 1 alpha 2; 7013; 7018; amplification; mammal; human; dog; cat;
 XX bile duct; colon; breast; uterus; oesophagus; larynx; liver; brain;
 XX remission; relapse; gene; ss.
 XX OS Synthetic.
 XX PN WO200198539-A2.
 XX PD 27-DEC-2001.
 XX PF 21-JUN-2001; 2001WO-US19980.
 XX PR 21-JUN-2000; 2000US-215727P.
 XX PR 27-OCT-2000; 2000US-243976P.

XX (HITB) HITACHI CHEM CO LTD.
 PA (HITB) HITACHI CHEM RES CENT INC.
 PA (HITA) HITACHI LTD.
 PI Mitsuhashi M, Kambara H, Matsunaga H, Kawamura M;
 XX MPI; 2002-098233/13.
 DR
 XX
 PT Identifying lung cancer/metastasis of solid tumor in patient by
 PT isolating blood or non-lung tissue, or bone marrow from patient and
 PT identifying presence of marker e.g. syndecan 1, collagen 1 alpha 2,
 PT 7013, or 7018 -
 XX
 PS Example 2, Fig 6; 29pp; English.
 PS
 XX This sequence corresponds to exons 2-6 of the Syndecan 1 gene.
 CC Syndecan 1 is a cell surface transmembrane heparan sulphate proteoglycan
 CC from the family of proteoglycans that bind to extracellular matrix and
 CC growth factors. The loss of regulation of this gene has been identified
 CC in several cancers. This sequence was identified using the method of
 CC the invention for identifying lung cancer or metastasis of a solid
 CC tumour. The method comprises isolating blood (or non-lung tissue in
 CC the case of identifying lung cancer, or bone marrow in case of
 CC identifying metastasis) from a patient, and identifying the presence
 CC of at least one marker (M) such as syndecan 1, collagen 1 alpha 2, 7013,
 CC or 7018. This sequence was found to be more abundant in lung cancer
 CC RNA than in normal blood. The method of the invention is useful for
 CC identifying lung cancer in a mammal e.g., human, dog or cat, and
 CC identifying metastasis of solid tumour in a patient, where the solid
 CC tumour is of bile duct, colon, breast, uterus, oesophagus or larynx. The
 CC method is useful for identifying presence of lung cancer cells in the
 CC blood or bone marrow, and also for identifying metastasis and thus for
 CC identifying lung cancer cells in an organ such as liver or brain. The
 CC method is useful to identify the presence of lung cancer cells at a
 CC very early stage in the disease, or after remission or to identify a
 CC relapse.
 CC
 XX
 SQ Sequence 425 BP; 62 A; 124 C; 146 G; 93 T; 0 other;
 Query Match 55.8%; Score 181.4; DB 24; Length 425;
 Best Local Similarity 94.5%; Pred. No. 1.9e-36;
 Matches 188; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 4 GAGGCGGCGAGGCGCTCGAGATCCTCTCGAGACAGCCCGCTGCTGCGCGCG 63
 Db 227 GAGGCGGCGAGGCGCTCGAGATCCTCTCGAGGCTCAAGCCCGCTGCTGCGCGCG 286
 QY 64 TCTCCAGGCGGCGCTGCTCTCTGAAATGACGAGGCGTCTGAGGCGAGCTGCT 123
 Db 287 TCTCCAGGCGGCGCTGCTCTCTGAAATGACGAGGCGTCTGAGGCGAGCTGCT 346
 QY 124 GAGCGCTTCATCCAGAGGCTCTCCGTAGCTCTGTGCGCCCACTGCGGCGCTG 183
 Db 347 GAGCGCTTCATCCAGAGGCTCTCCGTAGCTCTGTGCGCCCACTGCGGCGCTG 406
 QY 184 GCGTGAATCAGGAATATT 202
 Db 407 GCGTGAATCAGGAATATT 425

XX OS Homo sapiens.
 XX PN WO200204514-A2.
 XX PD 17-JAN-2002.
 XX PF 10-JUL-2001; 2001WO-US22058.
 XX PR 11-JUL-2000; 2000US-0614124.
 XX PR 29-AUG-2000; 2000US-0651563.
 XX PR 08-SEP-2000; 2000US-0658824.
 XX PR 26-SEP-2000; 2000US-0671325.
 XX PR 06-OCT-2000; 2000US-0677419.
 XX PR 30-OCT-2000; 2000US-0702705.
 XX PR 13-DEC-2000; 2000US-0736457.
 XX PR 03-MAY-2001; 2001US-0849626.
 XX PA (CORI-) CORIXA CORP.
 XX PI Wang T, Matanabe Y, Henderson RA, Johnson JC, Retter MW,
 XX PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS,
 XX PI Mcnabb A, Wang A, Fanger N, Switzer A, Monelli PD, Clapper JD;
 XX MPI; 2002-164634/21.
 DR
 XX
 PT Novel polynucleotide encoding a lung tumour polypeptide useful for
 PT stimulating and/or expanding T cells specific for a tumour protein -
 PS Claim 1; SEQ ID No 1041; 223pp; English.
 PS
 XX The invention describes an isolated polynucleotide and polypeptide
 CC useful for stimulating and/or expanding T cells specific for a tumour
 CC protein for determining the presence of a cancer in a patient. A
 CC composition containing the polynucleotide and/or polypeptide is useful
 CC for treating a lung cancer in a patient. The polypeptide is useful for
 CC removing tumour cells from a biological sample. The polynucleotide is
 CC also useful as probe or primer to detect the level of mRNA encoding a
 CC tumour protein. This sequence encodes a lung tumour associated protein
 CC or protein fragment, described in the method of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at
 CC ftp.wipo.int/pub/published_pcr_sequences.
 XX
 SQ Sequence 492 BP; 153 A; 126 C; 110 G; 102 T; 1 other;
 Query Match 44.4%; Score 144.4; DB 24; Length 492;
 Best Local Similarity 98.6%; Pred. No. 4.7e-27;
 Matches 145; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 162 TGTGCCCCACCCCTGGCGCCCTGGCTGAGATCAGAAATATTTCCAAAGATGATGCT 221
 Db 492 TGTGCCCCACCCCTGGCGCCCTGGCTGAGATCAGAAATATTTCCAAAGATGATGCT 433
 QY 222 TTTGCTTTTGGCAAACTCTACTTAATCCATGSGTTTCTCTGTGACAGTATTTCC 281
 Db 432 TTTGCTTTTGGCAAACTCTACTTAATCCATGSGTTTCTCTGTGACAGTATTTCC 373
 QY 282 AAATGTAATAAATCTTAATATAAGTA 308
 Db 372 AAATGTAATAAATCTTAATATAAGTA 346

RESULT 12
 AAK53003/c
 ID AAK53003 standard; cDNA; 492 BP.
 XX
 AC AAK53003;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE cDNA encoding lung tumour protein clone R0128.E06.
 XX
 KM Lung tumour; cancer; T cell; immune response stimulator;
 KM cytostatic; gene; ss.

RESULT 13
 AAK54184/c
 ID AAK54184 standard; cDNA; 286 BP.
 XX
 AC AAK54184;
 XX
 DT 16-NOV-2001 (first entry)
 XX
 DE Murine transport and binding associated protein encoding cDNA SEQ ID 749.


```

XX Murine; liver; gene library; amino acid synthesis; binding protein;
XX cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
XX phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
XX replication; transcription; translation; transport protein; ss.
XX Mus musculus.
OS DE20103510-U1.
XX 07-JUN-2001.
XX 28-FEB-2001; 2001DE-2003510.
XX 02-DEC-1999; 99DE-1058160.
XX (LION-) LION BIOSCIENCE AG.
XX WPI; 2001-368570/39.
XX
XX Gene library containing sequences with specific 3'-ends and no polyA
XX tail, encoding proteins involved in a wide range of cellular processes
XX
XX Claim 15; Page 227; 251pp; German.
XX
XX This invention describes a novel gene library (A) comprises a gene
XX sequence (or its part) encoding a protein involved in amino acid
XX synthesis, cellular/energy metabolism, metabolism of
XX fatty acids/phospholipids, synthesis or breakdown of
XX purines/pyrimidines/nucleosides/nucleotides; DNA
XX replication/transcription/translation, or, is a transport/binding protein.
XX (A) are produced that correspond to the 3'-end of mRNA but without the
XX polyA tail. They can be prepared more efficiently and with less effort
XX than conventional libraries. AAK53436-AAK54275 represent fragments of the
XX gene library described in the method of the invention.
XX
XX Sequence 286 BP; 78 A; 57 C; 74 G; 77 T; 0 other;
XX
XX Query Match 38.0%; Score 123.4; DB 22; Length 286;
XX Best Local Similarity 76.2%; Pred. No. 8e-22;
XX Matches 208; Conservative 0; Mismatches 51; Indels 14; Gaps 4;
XX
XX 36 GACCAAGCCGCTGCTGCTGAGCGCGCTCCAGGGGCTGCTTCTCTGGAATTGAC 95
XX 279 GCCCAAGCGCTCCACCTTGGTACCATCTCTAGTCACTCTCTCCGGAAGTTGAC 220
XX
XX 96 GAGGGGTGTCTGGGAGAGCTGCTGAGCGCCCTCCAGAGCGAGTTCCTCGT 155
XX 219 AAGACACATCTTGAGTATGCGTCACTG-GTTCCTCCATCAAGAACCAAGTTCACCTTC 161
XX
XX 156 AGCTCCGTGAGCCCAACCTCTGGGCTGGAATCAGAAATATTTTCCAAAGATGA 215
XX DB 160 AGCTCCGTGAGCCCAACCTCTGGGCTGGAATCAGAAATATTTTCCAAAGATGA 108
XX
XX 216 TAGCTTTTGTCTTTGGGAAACTCTAATTCACAAAGGTTTCTCTGACAGTAA 275
XX DB 107 -AGCTTTTGTCTTTGGGAAACTCTAATTCACAAAGGTTTCTCTGACAGTAA 54
XX
XX 276 TTTTCCAATGTAATTAACCTTAATTAATGA 308
XX DB 53 TTTTCCAATGTAATTAACCTTAATTAATGA 21
XX
XX RESULT 14
XX AA067902
XX ID AA067902 standard; DNA; 26700 BP.
XX AC AA067902;
XX XX
XX DT 08-DEC-1994 (first entry)
XX XX
XX DB Syndecan gene.

```

```

XX XX Syndecan; enhancer; differentiation; tumor; therapy; hair growth; ss.
XX XX
XX OS Mus sp.
XX
XX Key Location/Qualifiers
XX FT intron 1..4377
XX FT intron /tag= a
XX FT exon 4378..4443
XX FT intron /tag= b
XX FT intron 4444..22025
XX FT exon /tag= c
XX FT exon 22026..22106
XX FT intron /tag= d
XX FT intron 22107..23000
XX FT exon /tag= e
XX FT exon 23001..23483
XX FT intron /tag= f
XX FT intron 23484..23904
XX FT exon /tag= g
XX FT exon 23905..24039
XX FT intron /tag= h
XX FT intron 24040..24250
XX FT exon /tag= i
XX FT exon 24251..24418
XX FT intron /tag= j
XX FT intron 24419..26700
XX FT intron /tag= k
XX
XX M09412162-A.
XX PD 09-JUN-1994.
XX
XX 01-DEC-1993; 93WO-F100514.
XX PF
XX 01-DEC-1992; 92US-0988427.
XX PR
XX (WAER/) WAERRI A M.
XX PA (ALAN/) ALANEN-KURKI L M.
XX PA (AUVI/) AUVINEN P O V.
XX
XX Alanen-Kurki LM, Auvinen POV, Jaakkola PM, Jalankari MT;
XX Leppaesm, Mali MS, Vihinen TA, Waerri AM;
XX P-PSDB; AAR55276.
XX
XX WPI; 1994-199926/24.
XX
XX DR P-PSDB; AAR55276.
XX
XX Syndecan stimulation of cellular differentiation - useful for
XX decreasing tumor growth used to promote hair growth
XX
XX Disclosure; Page 22-39; 65pp; English.
XX
XX The mouse syndecan gene enhancer, located 8-10 kb upstream from the
XX initiation site, is given in AA067901. Manipulation of the enhancer
XX can be used either to slow or prevent tumor growth or to promote
XX differentiation of specific cell types, e.g. epidermal cells to
XX promote hair formation. The complete mouse syndecan gene and its
XX encoded protein are given in AA067902 and AAR55276.
XX
XX Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 1 other;
XX
XX Query Match 34.8%; Score 113; DB 15; Length 26700;
XX Best Local Similarity 76.2%; Pred. No. 1.6e-18;
XX Matches 208; Conservative 0; Mismatches 50; Indels 15; Gaps 5;
XX
XX 36 GACCAAGCCGCTGCTGCTGAGCGCGCTCCAGGGGCTGCTTCTCTGGAATTGAC 95
XX DB 25419 GCCCAAGCGCTCCACCTTGGTACCATCTTA-GTCACTCTCTCCGGAAGTTGAC 25477
XX
XX 96 GAGGGGTGTCTGGGAGAGCTGCTGAGCGCCCTCCAGAGCGAGTTCCTCGT 155
XX DB 25478 AAGACACATCTTGAGTATGCGTCACTG-GTTCCTCCATCAAGAACCAAGTTCACCTTC 25536

```

PR	07-MAR-1994;	94US-0206185.
PR	01-DEC-1992;	92US-0388427.
PR	01-DEC-1993;	93WO-FT00514.
XX	07-JUN-1995;	95US-0472217.
PA	(ALAN/) ALANEN-KURKI L.	
PA	(ADU/) AUVINEN P.	
PA	(JAAK/) JAAKKOLA P.	
PA	(JALK/) JALKANEN M.	
PA	(LEPP/) LEPPAS S.	
PA	(MALI/) MALI M.	
PA	(VIHI/) VIHINEN T.	
PA	(WAER/) WAERRI A.	
XX		
PI	Atanen-kurki L, Auvinen P, Jaakola P, Jalkanen M;	
PI	Leppae S, Mali M, Vihinen T, Waerri A;	
XX		
DR	WPI, 1998-192770/17.	
XX	P-PSDB; AAW47156.	
PT	New mouse syndecan gene sequences - useful for, e.g. suppressing	
PT	tumour growth or promoting tissue regeneration in processes such as	
PT	wound healing	
XX		
PS	Claim 2, Fig 2A-O; 48pp; English.	
XX		
CC	This is the mouse syndecan gene sequence. A 350 base pair fragment	
CC	(AAV15948) of a purified 236 base pair DNA molecule (AAV15947) enhances	
CC	the expression of a gene operably linked to the promoter of the mouse	
CC	syndecan gene in 3T3 cells following treatment with TGF- beta and bFGF	
CC	when the fragment is operably linked to the promoter. A purified DNA	
CC	molecule comprising a portion of the nucleotide residues 3538-3888 of	
CC	the mouse syndecan genomic sequence suppresses expression of a gene	
CC	operably linked to the promoter of the mouse syndecan gene in S115 cells	
CC	treated with testosterone. Host cells can be transfected with vectors	
CC	which contain either the enhancing or suppressing DNA molecules. The	
CC	products may be used to alter the differentiated state of a host cell by	
CC	altering its expression of syndecan, e.g. to induce and regulate	
CC	syndecan expression, especially in cells which exhibit a malignant	
CC	phenotype, regardless of the origin of transformation. The products can	
CC	be used to produce therapeutics for suppressing tumour growth. They may	
CC	enhance the syndecan expression in a host cell, by enhancing its gene	
CC	transcription, especially in malignant or normal cells, and therefore	
CC	promote tissue regeneration, especially in processes such as wound	
CC	healing.	
XX		
SQ	Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 1 other;	
Query Match	34.8%; Score 113; DB 19; Length 26700;	
Best Local Similarity	76.2%; Pred. No. 1.6e-18;	
Matches 208; Conservative	0; Mismatches 50; Indels 15; Gaps 5;	
OY	36 GACCAAGCCCTGCTGCTGTGGCGCGCTTCACAGG3GCTGCTTCTCTGGAATTGAC 95	
DB	25419 GCCACAGCGCTCTCCACCTTTGGTACATCTCTA-GTCACTCTCTCTCCGAGAGTGCAC 25477	
OY	96 GAGGGGTGTCTTGGGGCAGAGCTGCGCTGACGGCTCATCAAGCCAGGTTCCTCGTT 155	
DB	25478 AATACACATCTTGAGTATGSGT3GCACTG-GTCTTCTCATCAAGAAACAATTACCTTC 25536	
OY	156 AGCTCTGTGGCCCACTCGG3CCCTG3GCTGGAATCAGAGATATTTTCCAAAGATGA 215	
DB	25537 AGTCTCTGTGGCCCC-----GCCCCAGGCGCTGGAAGTCAGAAATGTCTTCCAAAGATG- 25589	
OY	216 TAGCTTTTGGCTTTGGCAAAACTACTATATCAATGGGTTTTTCTCTGTACAGTAGA 275	
DB	25590 -AGTCTTTTGGCTTTGGCAAAAGCTACTATATCAATGGGT-----TCTGTACAGTAGA 25643	
OY	276 TTTTCCAAATGATATAACTTAATTAACTA 308	
DB	25644 TTTGGAGTGTGATTAACCTTTAATATTAAGA 25676	

Fri Jan 10 09:34:16 2003

us-09-825-682a-57.rng

Page 11

Search completed: January 10, 2003, 01:54:11
Job time : 127.067 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:47:01 ; Search time 18.99 Seconds
(without alignments)
5248.554 Million cell updates/sec

Title: US-09-825-682A-57
Perfect score: 325
Sequence: 1 aaagagggcgagcgagcgct.....gtaaaaaaaaaaaaaaaa 325

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/6C_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB	ID Description
1	307.4	94.6 2430	US-08-488-199-3 Sequence 3, Appl1
2	285.8	87.9 4797	US-09-643-597-134 Sequence 134, App
3	113	34.8 26700	US-08-472-217-1 Sequence 1, Appl1
4	113	34.8 26700	US-08-488-199-5 Sequence 5, Appl1
5	113	34.8 26700	US-08-760-534A-1 Sequence 1, Appl1
6	112.6	34.6 2432	US-08-078-683A-1 Sequence 1, Appl1
7	38.8	11.9 958	US-08-757-046A-5 Sequence 5, Appl1
8	38.8	11.9 958	US-09-447-208-5 Sequence 5, Appl1
9	38.8	11.9 958	US-09-135-988-5 Sequence 5, Appl1
10	38.8	11.9 958	US-09-277-715-5 Sequence 5, Appl1
11	38.8	11.9 958	US-08-597-274A-5 Sequence 5, Appl1
12	38.8	11.9 958	US-08-908-909-5 Sequence 5, Appl1
13	38.8	11.9 958	US-09-609-161B-5 Sequence 5, Appl1
14	38.8	11.9 958	US-08-990-103-5 Sequence 5, Appl1
15	38.4	11.8 8920	US-08-446-855A-1 Sequence 1, Appl1
16	38.4	11.8 8920	US-09-150-741-1 Sequence 1, Appl1
17	38.2	11.8 1332	US-09-057-762-1 Sequence 1, Appl1
18	38.2	11.8 1332	US-08-326-119A-1 Sequence 1, Appl1
19	37.6	11.6 1260	US-08-599-252-79 Sequence 79, Appl1
20	37.6	11.6 1260	US-08-436-074-52 Sequence 52, Appl1
21	37.6	11.6 1260	PCT-US96-06352-79 Sequence 79, Appl1
22	37.6	11.6 1260	PCT-US96-06583-79 Sequence 79, Appl1
23	36.4	11.2 579	US-08-684-862-10 Sequence 10, Appl1
24	36.4	11.2 579	US-09-040-984-75 Sequence 75, Appl1
25	36.4	11.2 579	US-09-123-912-75 Sequence 75, Appl1
26	36.4	11.2 579	US-09-643-597-75 Sequence 75, Appl1
27	36.2	11.1 4257	US-08-690-473-1 Sequence 1, Appl1

28	36.2	11.1	4257	4	US-09-259-821A-1	Sequence 1, Appl1	
29	36.2	11.1	4257	4	US-08-843-659-1	Sequence 1, Appl1	
30	36.2	11.1	4337	3	US-09-187-049-1	Sequence 1, Appl1	
c	31	36.2	11.1	12001	1	US-08-458-568A-11	Sequence 11, Appl1
32	35.4	10.9	2851	4	US-09-535-521-1	Sequence 1, Appl1	
33	35.4	10.9	2851	4	US-09-535-521-3	Sequence 3, Appl1	
34	35.4	10.9	3785	1	US-08-445-640-9	Sequence 9, Appl1	
35	35.4	10.9	3785	3	US-08-170-558-9	Sequence 9, Appl1	
36	35.4	10.9	3785	3	US-08-447-314-9	Sequence 9, Appl1	
37	35.4	10.9	3785	3	US-08-445-461-9	Sequence 9, Appl1	
38	35.2	10.8	1172	1	US-07-945-288-9	Sequence 9, Appl1	
39	35.2	10.8	1172	1	US-08-462-831-9	Sequence 9, Appl1	
40	35.2	10.8	1172	1	US-08-461-809-9	Sequence 9, Appl1	
41	35.2	10.8	1172	1	US-08-461-441-9	Sequence 9, Appl1	
42	35.2	10.8	1172	5	PCT-US93-08518-9	Sequence 9, Appl1	
43	35.2	10.8	1841	5	PCT-US95-00362-1	Sequence 1, Appl1	
44	35	10.8	3214	1	US-08-484-105-17	Sequence 17, Appl1	
45	35	10.8	3214	1	US-08-484-106-17	Sequence 17, Appl1	

ALIGNMENTS

RESULT 1
US-08-488-199-3
; Sequence 3, Application US/08488199
; Patent No. 5851993
; GENERAL INFORMATION:
; APPLICANT: Valkeanen, Markku
; TITLE OF INVENTION: Suppression of Tumor Cell Growth By
; TITLE OF INVENTION: Syndecan-1 Ectodomain
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,199
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,862
; FILING DATE: 13-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbal, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1102.0130001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2430 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 206..1138
; US-08-488-199-3
Query Match 94.6%; Score 307.4; DB 2; Length 2430;
Best Local Similarity 99.1%; Pred. No. 1.7e-73;
Matches 319; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
QY      4  GAGGCGGAGGGGCTGGAGATCTCTCGACAGCACGCCGCTCGGCTGGGCGCG 63
      |||
Db      2100 GAGGCGGAGGGGCTGGAGATCTCTCGACAGCACGCCGCTCGGCTGG-6GCGG 2158
QY      64  TCTCCAGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTGTCTTGGGCAAGCTGGCTCT 123
      |||
Db      2159 TCTCCAGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTGTCTTGGGCAAGCTGGCTCT 2218
QY      124 GAGCGCTTCATCCAGGCGCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 183
      |||
Db      2219 GAGCGCTTCATCCAGGCGCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 2278
QY      184 GCGTGAATCGAGAAATATTTTCCAAAGAGTGAATGCTTTTGCTTTGGGAAAACCTTAC 243
      |||
Db      2279 GCGTGAATCGAGAAATATTTTCCAAAGAGTGAATGCTTTTGCTTTGGGAAAACCTTAC 2338
QY      244 TTATCCAAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 303
      |||
Db      2339 TTATCCAAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 2398
QY      304 AAGTAAAAAAAAAAAAAAAA 325
      |||
Db      2399 AAGTAAAAAAAAAAAAAAAA 2420

RESULT 2
US-09-643-597-134
; Sequence 134, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 134
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = A,T,C or G
US-09-643-597-134

Query Match      87.9%; Score 285.8; DB 4; Length 4797;
Best Local Similarity 96.1%; Pred. No. 1.4e-67;
Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
```

```
Db      3885 GAGCGCTTCATCCAAAGCGCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 3944
QY      184 GCGTGAATCGAGAAATATTTTCCAAAGAGTGAATGCTTTTGCTTTGGGAAAACCTTAC 243
      |||
Db      3945 GCGTGAATCGAGAAATATTTTCCAAAGAGTGAATGCTTTTGCTTTGGGAAAACCTTAC 4004
QY      244 TTATCCAAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 303
      |||
Db      4005 TTATCCAAATGGGTTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 4064
QY      304 AAGTA 308
      |||
Db      4065 AAGTA 4069

RESULT 3
US-08-472-217-1
; Sequence 1, Application US/08472217
; Patent No. 5726058
; GENERAL INFORMATION:
; APPLICANT: Alanen-Kurki, Leena
; APPLICANT: Auvinen, Petri
; APPLICANT: Jaakkola, Panu
; APPLICANT: Jalonen, Markku
; APPLICANT: Lepp, Sirpa
; APPLICANT: Maki, Markku
; APPLICANT: Vihtinen, Tapani
; APPLICANT: W rti, Anni
; TITLE OF INVENTION: Syndecan Stimulation Of Cellular
; TITLE OF INVENTION: Differentiation
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,217
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,186
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,427
; FILING DATE: 01-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1102.0050003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4378..4443, 22026..22106, 23001..23483,
; LOCATION: 23905..24039, 24251..24418)
```


RESULT 6
 US-08-078-683A-1
 Sequence 1, Application US/08078683A
 Patent No. 5486599
 GENERAL INFORMATION:
 APPLICANT: Saunders, Scott
 APPLICANT: Bernfield, Merton
 APPLICANT: Kato, Masato
 TITLE OF INVENTION: Construction and Use of Synthetic
 TITLE OF INVENTION: Constructs Encoding Syndecan
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 State Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII (text)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/078,683A
 FILING DATE: 17-JUN-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Vincent, Matthew P.
 REGISTRATION NUMBER: 36,709

Qy	36	GAACACAGCCCGCTCTGCGCTGGAGGCGCGCTTCAGAGGCGCTTCCTCTGAAATTGAC	95
Db	2176	GCCCAAGGCTCTCCACCTTTGGTAACATCTTA-GTCACTCTCTCTCCGGAAAGTTGAC	2234
Qy	96	GAGGGGCTGCTTGGGCAAGAGTGAGCTGTGACGCGCTTCATCCAGGCGAGGTTCTCCGTT	155
Db	2235	AAGACACATCTTTAGATAGATAGGCTGGCACTG-CTTCCTCCATCAAGAACCAAGTTACCTTC	2293
Qy	156	AAGCTCGATGAGCCCCACCCCTGAGGCGCCGGAGCTGAGATACAGAAATATTTTCGAAAGAGTA	215
Db	2294	AAGCTCTGTGTGCCCC-----GCCCCAGGCTGAGATGCAAAATGTTTCCCAAGAGATG-2344	
Qy	216	TAGCTTTTGTCTTTGSCAAAACTCTACTTAATCCAAATGGGTTTTTCTCTGTACAGTAGA	275
Db	2347	-AGTCTTTTCTTTTGGCAAAAGCGTACTTAATCAATGAGGT-----TCTGTACAGTAGA	2400
Qy	276	TTTTCGAAATGTAATAACTTTAATTAAG	306
Db	2401	TTTTGCAATGATATATTAATCTTAATTAAG	2431

RESULT 7
 US-08-757-046A-5
 : Sequence 5, Application US/08757046A
 : Patent No. 5876995
 : GENERAL INFORMATION:
 : APPLICANT: Bryan, Bruce
 : TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
 : NUMBER OF SEQUENCES: 14
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Brown, Martin, Haller & McClain
 : STREET: 1660 Union Street
 : CITY: San Diego
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 92101-2926
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FASTSEQ Version 1.5

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/757,046A
;; FILING DATE: 11-25-96
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/597,274
;; FILING DATE: 02-06-96
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L.
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 6680-105B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-238-0999
;; TELEFAX: 619-238-0062
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 958 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE:
;; ORIGINAL SOURCE:
;; FEATURE:
;; NAME/KEY: Coding Sequence
;; LOCATION: 115...702
;; OTHER INFORMATION: apoaeguorin-encoding gene
;; PUBLICATION INFORMATION:
;; AUTHORS: Inouye et al.
;; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
;; VOLUME: 92
;; PAGES: 3154-3158
;; DATE: (1985)
;; DOCUMENT NUMBER: PATENT NO.: 5,093,240
;; US-08-757-046A-5

Query Match 11.9%; Score 38.8; DB 2; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 212 GTGATAGCTTTTGGCAAACTCTACTTAATCCATGGGTTTCTCTGACAG 271
Db 812 GTGTGATTTTGTATAGGACAGATTAATCGATGATTAGTTGTTTTTAATCAA 871

Qy 272 TAGATTTTCCAAATGTAATAAATTATATTAAGTAAAAA 325
Db 872 CAGAACTTCAAAATCGAAAAAGTAAAAA 925

RESULT 8
; Sequence 5, Application US/09447208
; Patent No. 6113886
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Erman White & Mcauliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/447,208
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 0909/135,988
;; FILING DATE: 08-17-98
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/757,046
;; FILING DATE: 11-25-96
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/597,274
;; FILING DATE: 02-06-96
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L.
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 24727-105C
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-450-8400
;; TELEFAX: 619-450-8499
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 958 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE:
;; ORIGINAL SOURCE:
;; FEATURE:
;; NAME/KEY: Coding Sequence
;; LOCATION: 115...702
;; OTHER INFORMATION: apoaeguorin-encoding gene
;; PUBLICATION INFORMATION: PATENT NO.: 5,093,240
;; AUTHORS: Inouye et al.
;; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
;; VOLUME: 82
;; PAGES: 3154-3158
;; DATE: (1985)
;; US-09-447-208-5

Query Match 11.9%; Score 38.8; DB 3; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 212 GTGATAGCTTTTGGCAAACTCTACTTAATCCATGGGTTTCTCTGACAG 271
Db 812 GTGTGATTTTGTATAGGACAGATTAATCGATGATTAGTTGTTTTTAATCAA 871

Qy 272 TAGATTTTCCAAATGTAATAAATTATATTAAGTAAAAA 325
Db 872 CAGAACTTCAAAATCGAAAAAGTAAAAA 925

RESULT 9
; Sequence 5, Application US/09135988
; Patent No. 6152358
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Erman White & Mcauliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA

```

EARLIER APPLICATION NUMBER: 60/102,939
EARLIER FILING DATE: 1998-10-01
EARLIER APPLICATION NUMBER: 60/089,367
EARLIER FILING DATE: 1998-06-15
EARLIER APPLICATION NUMBER: 60/079,624
EARLIER FILING DATE: 1998-03-27
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 958
TYPE: DNA
ORGANISM: Aequorea (luminescent jellyfish)
FEATURE:
NAME/KEY: CDS
LOCATION: (115)..(702)
FEATURE:
OTHER INFORMATION: Apoaequorin-encoding gene
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: 5,093,240
PATENT FILING DATE: 1992-10-08
PUBLICATION DATE: 1992-03-03
PUBLICATION INFORMATION:
AUTHORS: Inouye, S.
TITLE: Cloning and sequence analysis of cDNA for the luminescent protein
JOURNAL: Proc. Natl. Acad. Sci. USA
VOLUME: 82(110)
PAGES: 3154-3158
DATE: 1985-05
US-09-277-716-5

Query Match          11.9%; Score 38.8; DB 4; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 212 GGTGAAGCTCTTTCCTTTGGCAAACTCTACTTAATCCATGAGGGTTTCTGTCAG 271
Db 812 GGTGTGATTTTGTAAATTAGGAAACGATTAAATGATGATGTTAGTGTGTTTTTAATCAA 871
QY 272 TAGATTTCCAAATGTATATACTTTAATATATAGTAAAAAAGTAAAAAAGTAAAAA 325
Db 872 CGAAGCTTACAAATCGAAAAAGTAAAAAAGTAAAAAAGTAAAAAAGTAAAAAAGTAAAAA 925

RESULT 11
US-08-597-274A-5
; Sequence 5, Application US/08597274A
; Patent No. 6247995
; GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
TITLE OF INVENTION: BIOLUMINESCENT NOVELTY ITEMS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,274A
FILING DATE: 02/06/96
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779

```

REFERENCE/DOCKET NUMBER: 6680-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 115...702
OTHER INFORMATION: apoaecuatorin-encoding gene
PUBLICATION INFORMATION:
DOCUMENT NUMBER: 5,093,240
AUTHORS: Inouye et al.
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 82
PAGES: 3154-3158
DATE: (1985)
US-08-597-274A-5

Query Match 11.9%; Score 38.8; DB 4; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 212 GGTATGCTTTTGGTTTGGCAAACTCTACTTATCCATGGGTTTCTCTGACG 271
DB 812 GGTGATTTTGTATTTAGGAACAGATTAAATCGAATGATTCTGTTTATATCA 871

QY 272 TAGATTTCCAAATGTAATAACTTTATATTAAGTAAAAA 325
DB 872 CAGAACTTCAAAATCGAAAAAGTAAAAA 925

RESULT 12
US-08-908-909-5
Sequence 5, Application US/08908909
Patent No. 6416960
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
TITLE OF INVENTION: DETECTION AND VISUALIZATION OF
TITLE OF INVENTION: NEOPLASTIC TISSUES AND OTHER TISSUES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FaetsSO Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/908,909
FILING DATE: 08-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/023,374
FILING DATE: 08-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6680-108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 115...702
OTHER INFORMATION: apoaecuatorin-encoding gene
PUBLICATION INFORMATION:
AUTHORS: Inouye et al.
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 82
PAGES: 3154-3158
DATE: (1985)
US-08-908-909-5

Query Match 11.9%; Score 38.8; DB 4; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 212 GGTATGCTTTTGGTTTGGCAAACTCTACTTATCCATGGGTTTCTCTGACG 271
DB 812 GGTGATTTTGTATTTAGGAACAGATTAAATCGAATGATTCTGTTTATATCA 871

QY 272 TAGATTTCCAAATGTAATAACTTTATATTAAGTAAAAA 325
DB 872 CAGAACTTCAAAATCGAAAAAGTAAAAA 925

RESULT 13
US-09-609-161B-5
Sequence 5, Application US/09609161B
Patent No. 643682
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Szent-Gyorgyi, Christopher
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS,
TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
FILE REFERENCE: 24729-121B
CURRENT APPLICATION NUMBER: US/09/609,161B
CURRENT FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 09/277,716
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/102,939
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/089,367
PRIOR FILING DATE: 1998-06-15
PRIOR APPLICATION NUMBER: 60/079,624
PRIOR FILING DATE: 1998-03-27
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 958
TYPE: DNA
ORGANISM: Aequorea (luminescent jellyfish)
FEATURE:
NAME/KEY: CDS

LOCATION: (115)..(702)
OTHER INFORMATION: Apoaequorin-encoding gene
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: US 07/105,602
PATENT FILING DATE: 1987-10-08
PUBLICATION DATE: 1992-03-03
PUBLICATION INFORMATION:
AUTHORS: Inouye, S.
AUTHORS: No. 6436682uchi, M.
AUTHORS: Sakaki, Y.
AUTHORS: Takagi, Y.
AUTHORS: Miyata, T.
AUTHORS: Iwanaga, S.
AUTHORS: Miyata, T.
AUTHORS: Tsuji, F.I.
TITLE: Cloning and sequence analysis of cDNA for the luminescent protein
TITLE: aequorin
JOURNAL: Proc. Natl. Acad. Sci. USA
VOLUME: 82(10)
PAGES: 3154-3158
DATE: 1985-05
US-09-609-161B-5

Query Match 11.9%; Score 38.8; DB 4; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 212 GTGATAGCTTTGCTTTGGCAAACTTACTTAATCCAAAGGGTTTCTCTGTACAG 271
DB 812 GTGTGATTTTGTATTTAGGACGATTAATCGATGATTTAGTTTATTCAA 871

QY 272 TAGATTTCCAAATGTATTAACCTTTAATATTAAGTAAAAA 325
DB 872 CAGAACTTCAATCGAAAAAGTAAAAA 925

RESULT 14
US-08-990-103-5
Sequence 5, Application US/08990103
Patent No. 6458547
GENERAL INFORMATION:
APPLICANT: Bruce J. Bryan
APPLICANT: Stephen Gaalema
APPLICANT: Randall B. Murphy
TITLE OF INVENTION: APPARATUS AND METHOD FOR DETECTING AND
TITLE OF INVENTION: IDENTIFYING INSECTICIOUS AGENTS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/990,103
FILING DATE: 12-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,675,
FILING DATE: 02-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,745
FILING DATE: 12-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6680-112

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 115..702
OTHER INFORMATION: apoaequorin-encoding gene
PUBLICATION INFORMATION:
AUTHORS: Inouye et al.
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 82
PAGES: 3154-3158
DATE: (1985)
DOCUMENT NUMBER: 5,093,240
US-08-990-103-5

Query Match 11.9%; Score 38.8; DB 4; Length 958;
Best Local Similarity 58.8%; Pred. No. 0.16;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 212 GTGATAGCTTTGCTTTGGCAAACTTACTTAATCCAAAGGGTTTCTCTGTACAG 271
DB 812 GTGTGATTTTGTATTTAGGACGATTAATCGATGATTTAGTTTATTCAA 871

QY 272 TAGATTTCCAAATGTATTAACCTTTAATATTAAGTAAAAA 325
DB 872 CAGAACTTCAATCGAAAAAGTAAAAA 925

RESULT 15
US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: O'Sullivan, William J
APPLICANT: Flores, Maria V
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderbye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
; INFORMATION FOR SRQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 11.8%; Score 38.4; DB 2; Length 8920;
Best Local Similarity 56.2%; Pred. No. 0.48;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 198 ATATTTTCCAAAGATGATGCTTTTGCTTTTGCGAAACTCTACTTATCCAAATGGGT 257
Db 8631 AATTTTGTATATATACAAATTTATTATTCACTCATATGTAACCAAAATGGTT 8690
QY 258 TTTTCTGTACAGTAGATTTTCCAAATGTAATAACTTAAATTAAGTAAGTAAAAA 317
Db 8691 TTTTCATTTACAAATTAATTTTAATAATTTAATAATTTAATAATTTAATAATTTA 8750
QY 318 AAAAAAA 325
Db 8751 AATATATA 8758

Search completed: January 10, 2003, 03:22:58
Job time : 58.99 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 10, 2003, 02:45:21 / Search time 74.3322 Seconds

(Without alignments)
1924.247 Million cell updates/sec

Title: US-09-825-682a-57

Perfect score: 325

Sequence: 1 aaagsggsggsggsggscct.....gtcaaaaaaaaaaaaaaa 325

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US05_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	291.4	89.7	1763	10	US-09-925-302-8
2	285.8	87.9	4797	10	US-09-735-705-134
3	285.8	87.9	4797	10	US-09-850-716A-134
4	285.8	87.9	4797	10	US-09-880-107-3339
5	285.8	87.9	4797	10	US-09-897-778-134
6	144.4	44.4	492	9	US-09-736-457-1041
7	144.4	44.4	492	9	US-09-902-941-1041
8	144.4	44.4	492	9	US-09-849-626-1041
9	120.4	37.0	253	9	US-10-079-623-76
10	44.6	13.7	748	10	US-09-910-943-361
11	42.8	13.2	617	10	US-09-764-877-121
12	41.6	12.8	375	10	US-09-960-352-13618
13	39.8	12.2	659158	9	US-09-771-208-20
14	39.6	12.2	277	10	US-09-960-352-12673
15	39.4	12.1	241	10	US-09-960-352-7904
16	38.8	11.9	289	10	US-09-880-107-1117
17	38.8	11.9	958	10	US-09-803-211-5
18	38.8	11.9	958	10	US-09-746-485A-5
19	38.6	11.9	215	10	US-09-960-352-5093

20	38.6	11.9	2455	10	US-09-918-909-25	Sequence 25, Appl
21	38.6	11.9	2509	10	US-09-925-301-540	Sequence 540, Appl
22	38.4	11.8	325	10	US-09-764-846-30	Sequence 30, Appl
23	38.4	11.8	621	10	US-09-764-846-105	Sequence 105, Appl
24	38.4	11.8	2000	9	US-09-938-842A-3307	Sequence 3307, Ap
25	38.2	11.8	1319	10	US-09-969-347-1179	Sequence 179, App
26	38	11.7	819	9	US-10-202-193-236	Sequence 236, App
27	38	11.7	3716	9	US-09-978-295A-210	Sequence 210, App
28	38	11.7	3716	9	US-09-978-697-210	Sequence 210, App
29	38	11.7	3716	9	US-09-978-832A-210	Sequence 210, App
30	38	11.7	3716	9	US-09-999-832A-210	Sequence 210, App
31	38	11.7	3716	9	US-09-978-189-210	Sequence 210, App
32	37.8	11.6	401	10	US-09-960-352-10503	Sequence 10503, A
33	37.8	11.6	1300	10	US-09-822-849A-3	Sequence 3, Appl1
34	37.6	11.6	291	10	US-09-960-352-1243	Sequence 1243, Ap
35	37.6	11.6	368	10	US-09-834-975-47	Sequence 47, Appl
36	37.4	11.5	102	10	US-09-998-598-1903	Sequence 1903, Ap
37	37.4	11.5	444	10	US-09-960-352-1281	Sequence 1281, Ap
38	37.4	11.5	1409	10	US-09-925-301-176	Sequence 176, Appl
39	37.4	11.5	1422	9	US-09-798-889-24	Sequence 24, Appl
40	37.4	11.5	2660	10	US-09-925-299-80	Sequence 80, Appl
41	37.2	11.4	2103	10	US-09-883-060-1	Sequence 1, Appl1
42	37	11.4	469	10	US-09-954-456-248	Sequence 248, App
43	37	11.4	469	10	US-09-954-456-476	Sequence 476, App
44	37	11.4	3110	10	US-09-764-877-3912	Sequence 3912, Ap
45	37	11.4	3110	10	US-09-764-877-3914	Sequence 3914, Ap

ALIGNMENTS

RESULT 1
US-09-925-302-8
Sequence 8, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P1104
CURRENT APPLICATION NUMBER: US/09/925.302
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124.270
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1763
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-302-8

Query Match 89.7%; Score 291.4; DB 10; Length 1763;

Best Local Similarity 99.3%; Pred. No. 3.2e-60; Matches 303; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	4	GAGGGCGGAGGGCGCTGAGATCTCTCGACACAGCCCTCTGCTGCGCGG	63
DB	752	GAGGGCGGAGGGCGCTGAGATCTCTCGACACAGCCCTCTGCTGCGCGG	811
QY	64	TCTCCAGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTCTTTGGGCGAGCTTGCT	123
DB	812	TCTCCAGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTCTTTGGGCGAGCTTGCT	871
QY	124	GAGCGCTTCATCCAGGCGAGAGTCTCCGTACTCTCTGAGCCCACTCGGCGCTG	183
DB	872	GAGCGCTTCATCCAGGCGAGAGTCTCCGTACTCTCTGAGCCCACTCGGCGCTG	930
QY	184	GCGTGAATCAGGAATATTTCCAAAGAGTATGCTTTTGGCTTTTGGCAAACTTAC	243
DB	931	GCGTGAATCAGGAATATTTCCAAAGAGTATGCTTTTGGCTTTTGGCAAACTTAC	990

QY	244	TTAATCCAAAGGGTTTTCTCTGACAGTGAATTTCCAAATGATATAAAGTTAATAAT	303
Db	991	TAAACCAATGGTTTTTCCCTGACAGTGAATTTCCAAATGATATAAAGTTAATAAT	1055
QY	304	AAAGTA	308
Db	1051	AAAGTA	1055

RESULT 2

```

US-09-735-705-134
Sequence 134, Application US/09735705
Patent No. US20020052329A1
GENERAL INFORMATION:
APPLICANT: Mang, Tonglong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Mang, Aijun
APPLICANT: Skeky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C.4
CURRENT APPLICATION NUMBER: US/09/735,705
CURRENT FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 419
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 134
LENGTH: 4797
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(4797)
OTHER INFORMATION: n = A,T,C or G
US-09-735-705-134

```

Query Match	87.9%;	Score 285.8;	DB 10;	Length 4797;
Best Local Similarity	96.1%;	Pred. No. 9.7e-59;		
Matches 293; Conservative	0;	Mismatches 12;	Indels 0;	Gaps 0;

OY	4	GAGGGCTGGCAAGGGAGCTGGAGATCTCTCTGACAGACAGCGCGTCTGGCTGTGGGCGCG	63
Db	3765	GAGGGCGGCAAGGGAGCTGGAGATCTCTCTGACAGCGTCTCTCTGTGGCGCG	3824
OY	64	TCTCAGGGGCTCTCTCTCTGGAAATTBACGAGGGGTGTCTTGGGCAAGACTGGGCTCT	123
Db	3825	TCTCAGGGGCTCTCTCTCTGGAAATTBACGAGGGGTGTCTTGGGCAAGACTGGGCTCT	3884
OY	124	GAGGCTCTGCATCCAGGCGCAGGTTCCTCGTAGTCTGTGGGCCCAACCTGGGGCCCTG	183
Db	3885	GAGGCTCTGCATCCAGGCGCAGGTTCCTCGTAGTCTGTGGGCCCAACCTGGGGCCCTG	3944
OY	184	GCGCTGGAATCAGAAATATTTTCCAAAGATGATGCTTTTGTGGCAAACTCTAC	243
Db	3945	GCGCTGGAATCAGAAATATTTTCCAAAGATGATGCTTTTGTGGCAAACTCTAC	4004
OY	244	TTAATCATGGGTTTTTCTGTGACGTGATTTTCCAATGTAATTAATTATATA	303
Db	4005	TTAATCATGGGTTTTTCTGTGACGTGATTTTCCAATGTAATTAATTATATA	4064
OY	304	AAAGTA 308	
Db	4065	AAAGTA 4069	

```

US-09-850-716A-134
/ Sequence 134, Application US/09850716A
/ Patent No. US2002011519A1
/ GENERAL INFORMATION:
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: McNeill, Patricia D.
/ APPLICANT: Retter, Marc W.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
/ FILE REFERENCE: 210121.455C15
/ CURRENT APPLICATION NUMBER: US/09/850,716A
/ NUMBER OF SEQ ID NOS: 440
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 134
/ LENGTH: 4797
/ TYPE: DNA
/ ORGANISM: Homo sapien
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)_(4797)
/ OTHER INFORMATION: n = A,T,C or G
US-09-850-716A-134

```

Query Match	87.9%;	Score 285.8;	DB 10;	Length 4797;
Best Local Similarity	96.1%;	Pred. No. 9.7e-59;		
Matches 293; Conservative	0;	Mismatches 12;	Indels 0;	Gaps 0

QY	4	GAGGGCGGCGAGGGGGCTGGAAATCTCTCTGCGAGACAGAGCCGCTCTGCTGATGGAGCGG	63
Db	3765	GAGGGCGGCGAGGGGGCTGGAAATCTCTCTGCGAGGTCAAGCCCGCTCTCTGTGGCGCGG	3824
QY	64	TCTCCAGGGGCTCTTCTCTCTGGAAATTTGACGAGGGGTCTTGGGCGAGGCTGGCTCT	123
Db	3825	TCTCCAGGGGCTCTTCTCTCTGGAAATTTGACGAGGGGTCTTGGGCGAGGCTGGCTCT	3884
QY	124	GAGGCGCTCCATCGAGGCGAGGTTCCTCGTAGGCTCTGTGGCCCGACCTGGGGCGCTG	183
Db	3885	GAGGCGCTCCATCGAGGCGAGGTTCCTCGTAGGCTCTGTGGCCCGACCTGGGGCGCTG	3944
QY	184	GCGTGGATCAGGAATATTTTCCAAAGGTGATAGTCTTTGCTTTGGCAAACTGTAC	243
Db	3945	GCGTGGATCAGGAATATTTTCCAAAGGTGATAGTCTTTGCTTTGGCAAACTGTAC	4004
QY	244	TTAATCCATGGGTTTTTCTCTGACAGAGATTTTCCAAATGATAATTAATTAAATAA	303
Db	4005	TTAATCCATGGGTTTTTCTCTGACAGATTTTCCAAATGATAATTAATTAAATAA	4064
QY	304	AAAGTA 308	
Db	4065	AAAGTA 4069	

RESULT 4

```

US-09-680-107-3939
/ Seq-ence 3939, Application US/09880107
/ Patent No. US020020142981A1
/ GENERAL INFORMATION:
/ APPLICANT: Horne, Darci T.
/ APPLICANT: Vockley, Joseph G.
/ APPLICANT: Scherf, Uwe
/ APPLICANT: Gene Logic, Inc.
/ TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
/ FILE REFERENCE: 44921-5028-WO
/ CURRENT APPLICATION NUMBER: US/09/880,107
/ CURRENT FILING DATE: 2001-06-14
/ PRIOR APPLICATION NUMBER: US 60/211,379
/ PRIOR FILING DATE: 2000-06-14
/ PRIOR APPLICATION NUMBER: US 60/237,054
/ PRIOR FILING DATE: 2000-10-02
/ NUMBER OF SEQ ID NOS: 3950
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 3939

```



```

; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z48199
; NAME/KEY: unsure
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3939

Query Match      87.9%; Score 285.8; DB 10; Length 4797;
Best Local Similarity 96.1%; Pred. No. 9.7e-59;
Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 4 GAGGGGGGAGGGGGCTGGAGATCCTCGACAGACAGCCCGCTCGCTGTGGGCGG 63
Db 3765 GAGGGGGGAGGGGGCTGGAGATCCTCGACAGGCTCAGCCCGTCTCTGTGGGCGG 3824

Qy 64 TCTCCAGGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTGTCTTGGGCAAGCTGCTCT 123
Db 3825 TCTCCAGGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTGTCTTGGGCAAGCTGCTCT 3884

Qy 124 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 183
Db 3885 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 3944

Qy 184 GGGTGAATCAGGAATATTTTCCAAAGAGTGAATGTTCTTTGGGCAAACTCTAC 243
Db 3945 GGGTGAATCAGGAATATTTTCCAAAGAGTGAATGTTCTTTGGGCAAACTCTAC 4004

Qy 244 TTATCCATAGGGTTTCTCTGTACAGTAGATTTTCCAAATGTATTAATCTTAATA 303
Db 4005 TTATCCATAGGGTTTCTCTGTACAGTAGATTTTCCAAATGTATTAATCTTAATA 4064

Qy 304 AAGTA 308
Db 4065 AAGTA 4069

RESULT 5
US-09-897-778-134
; Sequence 134, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Matherakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvik, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Matanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 135_501, 4421, 4467, 4468, 4658
; OTHER INFORMATION: n = A,T,C or G
US-09-897-778-134

Query Match      87.9%; Score 285.8; DB 10; Length 4797;
Best Local Similarity 96.1%; Pred. No. 9.7e-59;
Matches 293; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
```

```

Qy 4 GAGGGGGGAGGGGGCTGGAGATCCTCGACAGACAGCCCGCTCGCTGTGGGCGG 63
Db 3765 GAGGGGGGAGGGGGCTGGAGATCCTCGACAGGCTCAGCCCGTCTCTGTGGGCGG 3824

Qy 64 TCTCCAGGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTGTCTTGGGCAAGCTGCTCT 123
Db 3825 TCTCCAGGGGGCTGCTTCTCTCTGGAATTGACGAGGGGTGTCTTGGGCAAGCTGCTCT 3884

Qy 124 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 183
Db 3885 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTG 3944

Qy 184 GGGTGAATCAGGAATATTTTCCAAAGAGTGAATGTTCTTTGGGCAAACTCTAC 243
Db 3945 GGGTGAATCAGGAATATTTTCCAAAGAGTGAATGTTCTTTGGGCAAACTCTAC 4004

Qy 244 TTATCCATAGGGTTTCTCTGTACAGTAGATTTTCCAAATGTATTAATCTTAATA 303
Db 4005 TTATCCATAGGGTTTCTCTGTACAGTAGATTTTCCAAATGTATTAATCTTAATA 4064

Qy 304 AAGTA 308
Db 4065 AAGTA 4069

RESULT 6
US-09-736-457-1041/C
; Sequence 1041, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvik, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1041
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(492)
; OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1041

Query Match      44.4%; Score 144.4; DB 9; Length 492;
Best Local Similarity 98.6%; Pred. No. 1.5e-25;
Matches 145; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 162 TGTGGCCCACTCGGGCCCTCGGCTGGAATGAGATATTTTCCAAAGATGATAGTCT 221
Db 492 TGTGGCCCACTCGGGCCCTCGGCTGGAATGAGATATTTTCCAAAGATGATAGTCT 433

Qy 222 TTTGCTTTTGGCAAACTCTACTTAATCCAAATGGGTTTTTCTGTACAGTAGATTTTCC 281
Db 432 TTTGCTTTTGGCAAACTCTACTTAATCCAAATGGGTTTTTCTGTACAGTAGATTTTCC 373

Qy 282 AATGTAAATTAATCTTAATTAAGTA 308
Db 372 AATGTAAATTAATCTTAATTAAGTA 346
```

RESULT 7

```

US-09-902-941-1041/C
; Sequence 1041, Application US/09902941
; Patent No. US2002017952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Mamerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Ranger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1041
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 473
; OTHER INFORMATION: n = A,T,C or G
US-09-902-941-1041

```

Query Match 44.4%; Score 144.4; DB 9; Length 492;

Best Local Similarity 98.6%; Pred. No. 1.5e-25;

Matches 145; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 162 TGTGGCCCAACCTGTGGCCCTGTGGCTGGAATCAGAAATTTTCCAAAGATGATGCT 221
DB 492 TGTGGCCCAACCTGTGGCCCTGTGGCTGGAATCAGAAATTTTCCAAAGATGATGCT 433
QY 222 TTTGCTTTTGGCAAACTCTACTTAATCAATGAGTTTCTCTGTACAGTAGATTTTCC 281
DB 432 TTTGCTTTTGGCAAACTCTACTTAATCAATGAGTTTCTCTGTACAGTAGATTTTCC 373
QY 282 AAATGTATTAACCTTAATATAAGTA 308
DB 372 AAATGTATTAACCTTAATATAAGTA 346

```

RESULT 8

```

US-09-849-626-1041/C
; Sequence 1041, Application US/09849626
; Publication No. US2002019769A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Ranger, Gary
; APPLICANT: Wang, Tongtong
; APPLICANT: Wang, Ajjun
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1041
; LENGTH: 492

```

```

; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(492)
; OTHER INFORMATION: n = A,T,C or G
US-09-849-626-1041

```

Query Match 44.4%; Score 144.4; DB 9; Length 492;

Best Local Similarity 98.6%; Pred. No. 1.5e-25;

Matches 145; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 162 TGTGGCCCAACCTGTGGCCCTGTGGCTGGAATCAGAAATTTTCCAAAGATGATGCT 221
DB 492 TGTGGCCCAACCTGTGGCCCTGTGGCTGGAATCAGAAATTTTCCAAAGATGATGCT 433
QY 222 TTTGCTTTTGGCAAACTCTACTTAATCAATGAGTTTCTCTGTACAGTAGATTTTCC 281
DB 432 TTTGCTTTTGGCAAACTCTACTTAATCAATGAGTTTCTCTGTACAGTAGATTTTCC 373
QY 282 AAATGTATTAACCTTAATATAAGTA 308
DB 372 AAATGTATTAACCTTAATATAAGTA 346

```

RESULT 9

```

US-10-079-623-76
; Sequence 76, Application US/10079623
; Patent No. US20020169302A1
; GENERAL INFORMATION:
; APPLICANT: Hayukkala, Ilkka J.
; APPLICANT: Glenn, Matthew
; APPLICANT: Gligor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: mammary gland and methods for their use.
; FILE REFERENCE: 11000.1044C3
; CURRENT APPLICATION NUMBER: US/10/079,623
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 253
; TYPE: DNA
; ORGANISM: Bovine
US-10-079-623-76

```

Query Match 37.0%; Score 120.4; DB 9; Length 253;

Best Local Similarity 75.6%; Pred. No. 5.8e-20;

Matches 189; Conservative 0; Mismatches 56; Indels 5; Gaps 3;

```

QY 77 CTTCCTCTGGAATTTGACGAGGCGGTGCTTGGGACAGCTGCTGAGGCGCTGCATC 136
DB 4 CTTCCTCTGTAAGTACAGACGCTTGGGTACCTGGACCTGAGTGCCACG 63
QY 137 CAAGCCAGGTTCTCCCTTAAGCTCTGTGGCCCACTGGGCTGGATGAG 196
DB 64 GATGACAGGCTCACTTTATAGCT--CTGTAGCCTGACTGTGGCCGGAATCAG 121
QY 197 AATATTTTCCAAAGAGTATGCTTTTGTCTTTGGCAAAAC-TCCTATTATCAATGG 255
DB 122 AATA--TTCCAAAGAGTATAGTCTTTTGTCTTTGGCCAACTTTATTAATCCATGG 179
QY 256 GTTTTCTGTACAGTGAATTTCCAAATGATTAATTAATATAAGTAAAAA 315
DB 180 GTTTTCTGTACAGTGAATTTTCCAAATGATTAATTAATATAAGTATGCT 239
QY 316 AAAAAA 325
DB 240 GAAAAA 249

```

RESULT 10

US-09-910-943-361
 : Sequence 361, Application US/09910943
 : Patent No. US20020081610A1
 : GENERAL INFORMATION:
 : APPLICANT: Hemmati-Briandjou, Ali
 : TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
 : FILE REFERENCE: 7529/IG148US1
 : CURRENT APPLICATION NUMBER: US/09/910,943
 : CURRENT FILING DATE: 2001-07-23
 : NUMBER OF SEQ ID NOS: 742
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 361
 : LENGTH: 748
 : TYPE: DNA
 : ORGANISM: Xenopus laevis
 : FEATURE:
 : NAME/KEY: misc.feature
 : LOCATION: (1)..(748)
 : OTHER INFORMATION: n may be a or g or c or t/u
 US-09-910-943-361

Query Match 13.7%; Score 44.6; DB 10; Length 748;
 Best Local Similarity 58.8%; Pred. No. 0.079;
 Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 195 GGATATTTTCCAAAGAGTATAGTCTTTGCTTTGGCAAACTCTACTTAATCCATG 254
 Db 201 GAACCAATCTTAAAGCATGATGACCGCTGTAGTAAATCCCTATTAATATAT 260
 Qy 255 GGTTCCTCTGTACAGAGATTTTCCAAATGTAATTAACCTTAATTAAGTAAAAA 314
 Db 261 AGTCCCTCTATACAGTGTATATGTACTGTAATAAATGTTTGTAAAAAGAAAAA 320
 Qy 315 AAAAAAAAAA 325
 Db 321 AAAAAAAAAA 331

RESULT 11
 US-09-764-877-121/c
 : Sequence 121, Application US/09764877
 : Patent No. US20020147140A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosen et al;
 : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 : FILE REFERENCE: PC005
 : CURRENT APPLICATION NUMBER: US/09/764,877
 : CURRENT FILING DATE: 2001-01-17
 : Prior application data removed - refer to PALM or file wrapper
 : NUMBER OF SEQ ID NOS: 4031
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 121
 : LENGTH: 617
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: SITE
 : LOCATION: (317)
 : OTHER INFORMATION: n equals a,t,g, or c
 US-09-764-877-121

Query Match 13.2%; Score 42.8; DB 10; Length 617;
 Best Local Similarity 51.6%; Pred. No. 0.2;
 Matches 98; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 136 CCAAGCCAGTTCCTCCGTTAGCTCTGTGGCCCAACCTGGGCGCTGGAATCAG 195
 Db 196 CTAGTCAGCTACTCTCTGAGCTTCTCAATTTCCCAAGACCGAGATGAATGTGG 137
 Qy 196 GAATATTTTCCAAAGAGTATAGTCTTTGCTTTGGCAAACTCTACTTAATCCATG 255
 Db 136 CTGATCTTGGCCCAATCTTATGAGATTTCTTCTCCGCCACTGAGATTTGATG 77

Qy 256 GTTTTCCTGTGACAGTATGATTTTCCAAATGTAATTAATTAATTAAGTAAAAA 315
 Db 76 GTAGTTTCTTGAATTAATTAATTAATCAAGTCCAGCCCTGTAAAAA 17
 Qy 316 AAAAAAAAAA 325
 Db 16 AAAAAAAAAA 7

RESULT 12
 US-09-960-352-13618/c
 : Sequence 13618, Application US/09960352
 : Patent No. US20020137139A1
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Wesley C.
 : APPLICANT: Tao, Ningbing
 : APPLICANT: Byatt, John C.
 : APPLICANT: Mathialagan, Nagappan
 : TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 : FILE REFERENCE: 16511.006/37-21(10298)C
 : CURRENT APPLICATION NUMBER: US/09/960,352
 : CURRENT FILING DATE: 2001-09-24
 : NUMBER OF SEQ ID NOS: 15112
 : SEQ ID NO 13618
 : LENGTH: 375
 : TYPE: DNA
 : ORGANISM: Bos taurus
 : OTHER INFORMATION: Clone ID: 58-LIB3058-048-Q1-K1-G6
 US-09-960-352-13618

Query Match 12.8%; Score 41.6; DB 10; Length 375;
 Best Local Similarity 62.5%; Pred. No. 0.32;
 Matches 65; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 222 TTGCTTTGGCAAACTCTACTTAATCCATGAGTTTCTCTGACAGTGAATTTCC 281
 Db 175 TTGCTTTTACGATGTTTCTTTTAAACGGCCCTTTTCCACCGTTCCCATTAATAA 116
 Qy 282 AATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 325
 Db 115 ACAAGACAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72

RESULT 13
 US-09-771-208-20
 : Sequence 20, Application US/09771208
 : Patent No. US2002015564A1
 : GENERAL INFORMATION:
 : APPLICANT: MEDRANO, JUAN
 : APPLICANT: BRADFORD, ERIC
 : APPLICANT: HORVAT, SIMON
 : TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
 : FILE REFERENCE: 4077-923710US
 : CURRENT APPLICATION NUMBER: US/09/771,208
 : CURRENT FILING DATE: 2001-01-26
 : PRIOR APPLICATION NUMBER: US 08/999,477
 : PRIOR FILING DATE: 1997-12-29
 : NUMBER OF SEQ ID NOS: 20
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 20
 : LENGTH: 659158
 : TYPE: DNA
 : ORGANISM: Mus musculus
 : FEATURE:
 : NAME/KEY: misc.feature
 : LOCATION: (123459)..(123478)
 : OTHER INFORMATION: n is unidentified a, c, g, or t
 : NAME/KEY: misc.feature
 : LOCATION: (602465)..(602485)
 : OTHER INFORMATION: n is unidentified a, c, g, or t
 : NAME/KEY: misc.feature

```

; LOCATION: (546998)..(547017)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (494715)..(494814)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (390986)..(391005)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (346860)..(346823)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (317174)..(317193)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (280353)..(280373)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (271829)..(271848)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (183872)..(183891)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (170625)..(170645)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (132680)..(132700)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; OTHER INFORMATION: n is a, c, g, or t
US-09-771-208-20

```

```

Query Match 12.2%; Score 39.8; DB 9; Length 659158;
Best Local Similarity 47.1%; Pred. No. 11;
Matches 122; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```

```

QY 63 GTCTCCAGGGGCTCTTCTCTCTGGAATTGACGAGGGGTCTTGGCAGAGCTGCTC 122
DB 39253 GTTTTGAAGAGGTGTATATGAGACAGGCTCCAGGGGTGGGGGTTAAAGCTCG 39312
QY 123 TGACGCGCTCCATCCAGGCCAGGTCCTCGTTAGCTCTGGGCCCCACCTGGGCCCT 182
DB 39313 TGAAGTGAGCTTTGGCAAGTGGCTCCGAGTAAAGTCTTCTCTAAGCTGATGACT 39372
QY 183 GGGCTGGATCAGCAATATTTTCCAAAGATGATAGTCTTTTGGCAAACTCTA 242
DB 39373 GAGTTGATCCCGAATCCATGTGTGGAAGAGAGACTACTGTCTCTGGCCCCCA 39432
QY 243 CTATATCAATGGGTTTCTCTGTACGTAGATTTTCCAAATGTAATAACTTAAAT 302
DB 39433 CATGGGACCCAGGGTACTGAGTGTACAGCCCATGTCCTCAGAAAGAAAGTCAATG 39492
QY 303 AAAGTAAAAAATAAAAAA 321
DB 39493 AATTAAAAAATAAAAAA 39511

```

```

RESULT 14
US-09-960-352-12673/c
; Sequence 12673, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112

```

```

; SEQ ID NO 12673
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 54-LIB3058-039-Q1-K1-P10
US-09-960-352-12673

```

```

Query Match 12.2%; Score 39.6; DB 10; Length 277;
Best Local Similarity 56.0%; Pred. No. 0.86;
Matches 75; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

```

QY 192 TCAGAAATATTTTCCAAAGATGATGCTTTGGTGGAAACTCTACTAATCCA 251
DB 219 TCAGAAATATTTTTCATTTTAAAAATTTTCTTTTAAATTTTAAATTTT 160
QY 252 ATGGGTTTCTCTGTACGTAGATTTTCCAAATGTAATTAATTAATTAATTAAT 311
DB 159 TTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 100
QY 312 AAAAAAAAAAAAAA 325
DB 99 AAAAAAAAAAAAAA 86

```

```

RESULT 15
US-09-960-352-7904/c
; Sequence 7904, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7904
; LENGTH: 241
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB3057-019-Q1-K1-A10
US-09-960-352-7904

```

```

Query Match 12.1%; Score 39.4; DB 10; Length 241;
Best Local Similarity 67.9%; Pred. No. 0.91;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```

```

QY 245 TAATCAATGGGTTTCTCTGTACGTAGATTTTCCAAATGTAATTAATTAATTA 304
DB 211 TAATCAATGATTAATTAATTTTAAAAAATTTTAAATTAATTAATTAATTA 152
QY 305 AGTAAAAAATAAAAAA 325
DB 151 AATAAAAAATAAAAAA 131

```

```

Search completed: January 10, 2003, 04:50:01
Job time : 196.332 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:46:01 ; Search time 697.746 Seconds
(without alignments)
7543.621 Million cell updates/sec

Title: US-09-825-682a-57

Sequence: 1 aaagagggcgagggcgct.....gtaaaaaaaaaaaaaa 325

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	ID	Description
1	320.4	98.6	513	9	AI224154 q135f10.x
2	320.4	98.6	534	14	BQ644995 AGENCOURT
3	320.4	98.6	722	14	BM980525 UI-CF-EN1
4	320.4	98.6	722	14	BM982212 UI-CF-EN1
5	318.4	98.0	327	14	BQ023686 UI-1-BB0-
6	318.4	98.0	330	14	BQ023676 UI-1-BB0-

C	7	317.4	97.7	615	9	AI831036	AI831036 wj62a07.x
C	8	316.8	97.5	666	14	BQ188527	BQ188527 UI-E-EJ1-
C	9	315.8	97.2	614	10	AM337157	AM337157 xw81f06.x
C	10	315.6	97.1	367	9	AI872681	AI872681 w08a02.x
C	11	315.6	97.1	590	14	BM968687	BM968687 UI-CF-DU1
C	12	314.4	96.7	548	9	AI224622	AI224622 qw96h04.x
C	13	314	96.6	519	10	AM192078	AM192078 x179f03.x
C	14	313.6	96.5	461	10	AM662594	AM662594 h133a10.x
C	15	312.4	96.1	636	9	AI691025	AI691025 tq13e11.x
C	16	311.4	95.8	483	10	BE206146	BE206146 b698f05.x
C	17	311.4	95.8	501	10	AM182758	AM182758 xp97c06.x
C	18	310.4	95.5	619	10	AM170078	AM170078 xm65b06.x
C	19	309.6	95.3	495	9	AA632080	AA632080 np70g10.s
C	20	309.4	95.2	721	14	BQ772055	BQ772055 UI-H-EJ1-
C	21	309.2	95.1	635	9	AI829106	AI829106 wj38g05.x
C	22	307.6	94.7	960	14	BQ955494	BQ955494 AGENCOURT
C	23	307.6	94.6	445	14	BM695154	BM695154 UI-E-CQ1-
C	24	307.6	94.6	449	13	BM666332	BM666332 UI-E-CQ1-
C	25	307.6	94.6	691	14	BQ774844	BQ774844 UI-H-FH0-
C	26	306.8	94.4	647	9	AA526849	AA526849 n192e04.s
C	27	306.4	94.3	562	9	AI566209	AI566209 tq69g09.x
C	28	306.4	94.3	592	9	AI921282	AI921282 w022g11.x
C	29	306.4	94.3	594	9	AI831457	AI831457 wj49b08.x
C	30	305.8	94.1	376	10	AM050972	AM050972 w225f01.x
C	31	305.8	94.1	443	14	BM683987	BM683987 UI-E-EJ1-
C	32	305.4	94.0	508	9	AA723113	AA723113 z980a06.s
C	33	305.4	94.0	624	10	AM439517	AM439517 x119f02.x
C	34	304.4	93.7	440	9	AI342919	AI342919 q036e09.x
C	35	304.4	93.7	548	9	AI677918	AI677918 w08e01.x
C	36	304.4	93.7	740	9	AI696816	AI696816 w074c09.x
C	37	303.4	93.4	348	10	AM273894	AM273894 xv73g01.x
C	38	303.4	93.4	409	14	BQ026547	BQ026547 UI-1-BB0-
C	39	303.4	93.4	422	9	AA426099	AA426099 zv47d04.s
C	40	303.4	93.4	464	9	AI078484	AI078484 o213g03.x
C	41	303.4	93.4	479	9	AI188950	AI188950 q022a04.x
C	42	303.4	93.4	484	12	BE697490	BE697490 RC0-CT042
C	43	303.4	93.4	490	9	AI148519	AI148519 q062c09.x
C	44	303.4	93.4	569	12	BE767978	BE767978 qv1-GN006
C	45	303.4	93.4	594	14	BQ026339	BQ026339 UI-1-BB0-

ALIGNMENTS

RESULT 1
LOCUS AI224154/1 513 bp mRNA linear EST 28-JAN-1999
DEFINITION g135f10.x1 Soares NHMPPU SI Homo sapiens cDNA clone IMAGE:1858507
3' similar to gb.J05392 SYNDECAN-1 PRECURSOR (HUMAN);, mRNA.

ACCESSION AI224154
VERSION AI224154.1 GI:3806867
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS (Bases 1 to 513)
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 834 Std Error: 0.00

Seq primer: -40UP from Gibco
High quality sequence stop: 322.
Location/Qualifiers

FEATURES
source

1. 513
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="IMAGE:1858507"
/clone_lib="Soares NIHMPU S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not 1;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBH, pregnant uterus
NBHPU, and fetal heart NBH119H) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of T.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 117 a 146 c 137 g 110 t 3 others
ORIGIN

Query Match 98.6%; Score 320.4; DB 9; Length 513;
Best Local Similarity 99.7%; Pred. No. 2.4e-42;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAGGGGCGAGGGGCTCGAGATCTCTCGACACACGCCCTGCTGCGGCGCG 63
DB 323 GAGGGGCGAGGGGCTCGAGATCTCTCGACACACGCCCTGCTGCGGCGCG 264
QY 64 TCTCCAGGGGCTGCTCTCTCTGAAATTGACGAGGGGTGCTTGGGCGAGCTGCT 123
DB 263 TCTCCAGGGGCTGCTCTCTCTGAAATTGACGAGGGGTGCTTGGGCGAGCTGCT 204
QY 124 GAGCGCTTCATCCAGGCGAGGTTCTCGGTAGCTCTGTGCGCCACCTGGGCGCTG 183
DB 203 GAGCGCTTCATCCAGGCGAGGTTCTCGGTAGCTCTGTGCGCCACCTGGGCGCTG 144
QY 184 GCGTGAATCAGGAATTTTCCAAAGAGTAGTCTTTGGTGGCAAACCTAC 243
DB 143 GCGTGAATCAGGAATTTTCCAAAGAGTAGTCTTTGGTGGCAAACCTAC 84
QY 244 TTAATCAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATA 303
DB 83 TTAATCAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATA 24
QY 304 AAGTAAAAAAAAAAAAAAAA 325
DB 23 AAGTAAAAAAAAAAAAAAAA 2

RESULT 2
B0644995 534 bp mRNA linear EST 15-JUL-2002
LOCUS B0644995
DEFINITION AGENCOURT_8498888 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6297772
5', mRNA sequence.
ACCESSION B0644995
VERSION B0644995.1 GI:21769167
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 534)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LHCW2507 row: m column: 05
High quality sequence stop: 532.
Location/Qualifiers
1. 534
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6297772"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGACAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 110 a 137 c 153 g 126 t 8 others
ORIGIN

Query Match 98.6%; Score 320.4; DB 14; Length 534;
Best Local Similarity 99.7%; Pred. No. 2.4e-42;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAGGGGCGAGGGGCTCGAGATCTCTCGACACACGCCCTGCTGCGGCGCG 63
DB 203 GAGGGGCGAGGGGCTCGAGATCTCTCGACACACGCCCTGCTGCGGCGCG 262
QY 64 TCTCCAGGGGCTGCTCTCTCTGAAATTGACGAGGGGTGCTTGGGCGAGCTGCT 123
DB 263 TCTCCAGGGGCTGCTCTCTCTGAAATTGACGAGGGGTGCTTGGGCGAGCTGCT 322
QY 124 GAGCGCTTCATCCAGGCGAGGTTCTCGGTAGCTCTGTGCGCCACCTGGGCGCTG 183
DB 323 GAGCGCTTCATCCAGGCGAGGTTCTCGGTAGCTCTGTGCGCCACCTGGGCGCTG 382
QY 184 GCGTGAATCAGGAATTTTCCAAAGAGTAGTCTTTGGTGGCAAACCTAC 243
DB 383 GCGTGAATCAGGAATTTTCCAAAGAGTAGTCTTTGGTGGCAAACCTAC 442
QY 244 TTAATCAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATA 303
DB 443 TTAATCAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATA 502
QY 304 AAGTAAAAAAAAAAAAAAAA 325
DB 503 AAGTAAAAAAAAAAAAAAAA 524

RESULT 3
BM980525/c 722 bp mRNA linear EST 21-MAR-2002
LOCUS BM980525
DEFINITION UI-CF-EN1-add-c-24-0-UI-s1 UI-CF-EN1 Homo sapiens cDNA clone
UI-CF-EN1-add-c-24-0-UI 3', mRNA sequence.
ACCESSION BM980525
VERSION BM980525.1 GI:19602077
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 722)
Normalizaton and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this CDNA
 sequence: 1-39, >AT-richflow_complexity (matched complement)
 617-637, >AT-richflow_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..722
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-CF-EN1-add-c-24-0-UI"
 /clone_lib="UI-CF-EN1"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial
 Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: Ecor I; Site 2: Not I;
 UI-CF-EN1 is a normalized CDNA library containing the
 following tissue(s): Primary Lung Cystic Fibrosis
 Epithelial Cells. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an Ecor I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CTCCTCAGGT.
 TAG_LIB=UI-CF-EN1
 TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
 6hr to LPS 24h
 TAG_SEQ=CTGCTCAGGT"
 BASE COUNT 195 a 201 c 170 g 155 t 1 others
 ORIGIN

Query Match

Best Local Similarity 99.7%; Score 320.4; DB 14; Length 722;
 Pred. No. 2e-42; Mismatches 1; Indels 0; Gaps 0;

Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

4 GAGGGGCGAGGGGCTGAGATCTCTGCGAGACCGCCCTGCTGCGGGCGG 63
 322 GAGGGGCGAGGGGCTGAGATCTCTGCGAGACCGCCCTGCTGCGGGCGG 263
 64 TCTCCAGGGGCTGCTCTCTCTGGAATTGACGAGGGGCTGCGGAGAGCTGCT 123
 262 TCTCCAGGGGCTGCTCTCTCTGGAATTGACGAGGGGCTGCGGAGAGCTGCT 203
 124 GAGCGCTTCATCCAGGCGCAGGTTCTCCGTTAGCTCTGTGCGCCCACTGCGGCCCTG 183
 202 GAGCGCTTCATCCAGGCGCAGGTTCTCCGTTAGCTCTGTGCGCCCACTGCGGCCCTG 143
 184 GGTGGAATCAGGAATATTTTCCAAAGAGTGATGCTTTTGGCTTTGGCAAAAGCTTAC 243
 142 GGTGGAATCAGGAATATTTTCCAAAGAGTGATGCTTTTGGCTTTGGCAAAAGCTTAC 83
 244 TTATTCATAGGATTTTCTCTGACAGTAGATTTTCCAAATGTAATTAATTAATA 303
 82 TTATTCATAGGATTTTCTCTGACAGTAGATTTTCCAAATGTAATTAATTAATA 23
 304 AAGTAAAAAATTTTCTCTGACAGTAGATTTTCCAAATGTAATTAATTAATA 325
 22 AAGTAAAAAATTTTCTCTGACAGTAGATTTTCCAAATGTAATTAATTAATA 1

RESULT 4
 BM982212/c
 LOCUS
 DEFINITION
 UI-CF-EN1-acr-o-03-0-UI.s1 UI-CF-EN1 Homo sapiens CDNA clone
 UI-CF-EN1-acr-o-03-0-UI 3', mRNA sequence.
 ACCESSION
 BM982212
 VERSION
 BM982212.1 GI:19605484
 EST

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 722)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

Query Match 98.6%; Score 320.4; DB 14; Length 722;
 Best Local Similarity 99.7%; Pred. No. 2e-42;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

4 GAGGGGCGAGGGGCTGAGATCTCTGCGAGACCGCCCTGCTGCGGGCGG 63

```

Db      322 GAGGCGGCGAGGCGCTGAGATCCTCCTGACGACACGCGCCGCTCCTGCTGCGCGCG 263
QY      64 TCTCCAGGGGCTGCTTCTCTCTGGAATTGAAGAGGGGTCTTTGGGGGAGACTGCTCT 123
Db      262 TCTCCAGGGGCTGCTTCTCTCTGGAATTGAAGAGGGGTCTTTGGGGGAGACTGCTCT 203
QY      124 GAGCGCTCCATCCAGGCGCAGGATCTCCGTTAGCTCTCTGCGCCCACTCGGCGCGCTG 183
Db      202 GAGCGCTCCATCCAGGCGCAGGATCTCTCGTTAGCTCTCTGCGCCCACTCGGCGCGCTG 143
QY      184 GCGTGGATCGAGATATTTTCCAAAGAGTATGCTTTTGGTCTTTGGCAAACTCTAC 243
Db      142 GCGTGGATCGAGATATTTTCCAAAGAGTATGCTTTTGGTCTTTGGCAAACTCTAC 83
QY      244 TTAATCCATGGGTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 303
Db      82 TTAATCCATGGGTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 23
QY      304 AAGTAAAAAAGGTTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 325
Db      22 AAGTAAAAAAGGTTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 1

RESULT 5
B0023686/c      327 bp      mRNA      linear      EST 27-MAR-2002
LOCUS      B0023686
DEFINITION      UI-1-BB0-abv-g-06-0-UI.s1 NCI CGAP P14 Homo sapiens cDNA clone
ACCESSION      B0023686
VERSION      B0023686.1 GI:19758965
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-38, >AT rich#low_complexity
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source      1..327
              Location/Qualifiers
              1..327
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="UI-1-BB0-abv-g-06-0-UI"
              /clone_lib="NCI CGAP P14"
              /tissue_type="Placenta"
              /dev_stage="Full Term"
              /lab_host="DH10B (Life Technologies)"
              /note="Organ: Placenta; Vector: pRT3-Pac (Pharmacia) with
              a modified polylinker; Site 1: EcoR I; Site 2: Not I;
              NCI CGAP P14 is a cDNA library containing the following
              tissue(s): Placenta full term. The library was constructed
              according to Bonaldo, Lennon and Soares, Genome Research,
              6:791-806, 1996. First strand cDNA synthesis was primed
              with an oligo-dT primer containing a Not I site. Double
              stranded cDNA was ligated to an EcoR I adaptor, digested
              with Not I, and cloned directionally into pRT3-Pac
              vector. The oligonucleotide used to prime the synthesis of
              first-strand cDNA contains a library tag sequence that is
              located between the Not I site and the (GT)18 tail. The

```

```

sequence tag for this library is AGGAA.
TAG_LIB=UI-1-BB0
TAG_TISSUE=Placenta human full term
TAG_SEQ=AGGAA"
BASE COUNT      83 a      84 c      80 g      79 t      1 others
ORIGIN
Query Match      98.0%; Score 318.4; DB 14; Length 327;
Best Local Similarity 99.4%; Fred. No. 6.6e-42;
Matches 319; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
4 GAGGCGGCGAGGCGCTGAGATCCTCCTGACGACACGCGCCGCTCCTGCTGCGCGCG 63
Db      321 GAGGCGGCGAGGCGCTGAGATCCTCCTGACGACACGCGCCGCTCCTGCTGCGCGCG 262
QY      64 TCTCCAGGGGCTGCTTCTCTCTGGAATTGAAGAGGGGTCTTTGGGGGAGACTGCTCT 123
Db      261 TCTCCAGGGGCTGCTTCTCTCTGGAATTGAAGAGGGGTCTTTGGGGGAGACTGCTCT 202
QY      124 GAGCGCTCCATCCAGGCGCAGGATCTCCGTTAGCTCTCTGCGCCCACTCGGCGCGCTG 183
Db      201 GAGCGCTCCATCCAGGCGCAGGATCTCTCGTTAGCTCTCTGCGCCCACTCGGCGCGCTG 142
QY      184 GCGTGGATCGAGATATTTTCCAAAGAGTATGCTTTTGGTCTTTGGCAAACTCTAC 243
Db      141 GCGTGGATCGAGATATTTTCCAAAGAGTATGCTTTTGGTCTTTGGCAAACTCTAC 82
QY      244 TTAATCCATGGGTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 303
Db      81 TTAATCCATGGGTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 22
QY      304 AAGTAAAAAAGGTTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 324
Db      21 AAGTAAAAAAGGTTTTTCTCTGTCAGTATGTTTCCAAATGTATTAATTAATA 1

RESULT 6
B0023676/c      330 bp      mRNA      linear      EST 27-MAR-2002
LOCUS      B0023676
DEFINITION      UI-1-BB0-abv-f-07-0-UI.s1 NCI CGAP P14 Homo sapiens cDNA clone
ACCESSION      B0023676
VERSION      B0023676.1 GI:19758955
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-38, >AT rich#low_complexity
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source      1..330
              Location/Qualifiers
              1..330
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="UI-1-BB0-abv-f-07-0-UI"
              /clone_lib="NCI CGAP P14"
              /tissue_type="Placenta"
              /dev_stage="Full Term"

```



```

/lab_host="DH10B (Life Technologies)"
/notes="Organ: Placenta; Vector: pT73-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI-CGAP_P14 is a cDNA library containing the following
tissue(s): Placenta full term. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is AGGA.
TAG_LIB=UI-1-B80
TAG_TISSUE=Placenta human full term
TAG_SEQ=AGGA"

BASE COUNT      83 a      85 c      81 g      80 t      1 others
ORIGIN

Query Match      98.0%; Score 318.4; DB 14; Length 330;
Best Local Similarity 99.4%; Pred. No. 6.6e-42;
Matches 319; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4  GAGGCGGCGAGGGGCTGGAGATCTCTCTGACAGCCAGCCGCTGCGGCGCG 63
DB      321 GAGGCGGCGAGGGGCTGGAGATCTCTCTGACAGCCAGCCGCTGCGGCGCG 262

QY      64  TCTCCAGGGGGCTGCTCTCTCTGAAATTCAGAGGGGTGTCTTGAGCAGAGCTGCT 123
DB      261 TCTCCAGGGGGCTGCTCTCTCTGAAATTCAGAGGGGTGTCTTGAGCAGAGCTGCT 202

QY      124 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTTG 183
DB      201 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTTG 142

QY      184 GCGTGAATCAGGAATATTTTCCAAAGAGTGAATAGTCTTTGGCTTTGGGAAAATCTTAC 243
DB      141 GCGTGAATCAGGAATATTTTCCAAAGAGTGAATAGTCTTTGGCTTTGGGAAAATCTTAC 82

QY      244 TTAATCCATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 303
DB      81  TTAATCCATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 22

QY      304 AAGTAAAAAAAAAAAAAAAA 324
DB      21  AAGTAAAAAAAAAAAAAAAA 1

RESULT 7
A1831036/c      615 bp      mRNA      linear      EST 21-DEC-1999
LOCUS      w162a07.x1 NCI CGAP Lul9 Homo sapiens cDNA clone IMAGS:2407380 3'
DEFINITION      similar to gb:J05392 SYNDSCAN-1 PRCURSOR (HUMAN), contains PMS.13
PMS1 repetitive element ;, mRNA sequence.

ACCESSION      A1831036
VERSION      A1831036.1 GI:5451707
KEYWORDS      EST.

SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 615)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

```

```

DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/Link at:
www.bio.lnlnl.gov/bdrp/image/image.html
Insert Length: 1533 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence crop: 443.
Location/Qualifiers
1. 615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2407380"
/clone_1ib="NCI-CGAP_Lul9"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT      156 a      179 c      153 g      127 t
ORIGIN

Query Match      97.7%; Score 317.4; DB 9; Length 615;
Best Local Similarity 99.7%; Pred. No. 6.6e-42;
Matches 318; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4  GAGGCGGCGAGGGGCTGGAGATCTCTCTGACAGCCAGCCGCTGCGGCGCG 63
DB      319 GAGGCGGCGAGGGGCTGGAGATCTCTCTGACAGCCAGCCGCTGCGGCGCG 260

QY      64  TCTCCAGGGGGCTGCTCTCTCTGAAATTCAGAGGGGTGTCTTGAGCAGAGCTGCT 123
DB      259 TCTCCAGGGGGCTGCTCTCTCTGAAATTCAGAGGGGTGTCTTGAGCAGAGCTGCT 200

QY      124 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTTG 183
DB      199 GAGCGCTTCATCCAGGCCAGGTTCTCCGTTAGCTCTGTGGGCCCACTGGGCGCTTG 140

QY      184 GCGTGAATCAGGAATATTTTCCAAAGAGTGAATAGTCTTTGGCTTTGGGAAAATCTTAC 243
DB      139 GCGTGAATCAGGAATATTTTCCAAAGAGTGAATAGTCTTTGGCTTTGGGAAAATCTTAC 80

QY      244 TTAATCCATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 303
DB      79  TTAATCCATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATACTTAATATA 20

QY      304 AAGTAAAAAAAAAAAAAAAA 322
DB      19  AAGTAAAAAAAAAAAAAAAA 1

RESULT 8
BQ188527
LOCUS      BQ188527
DEFINITION      UI-E-EJ1-a3w-j-12-0-UI-r1 UI-E-EJ1 Homo sapiens cDNA clone
UI-E-EJ1-a3w-j-12-0-UI 5', mRNA sequence.

ACCESSION      BQ188527
VERSION      BQ188527.1 GI:20364078
KEYWORDS      EST.

SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 666)
AUTHORS      Bonaldo M.F., Lennon, G. and Soares, M.B.

```

TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704447
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iue.wesg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA
sequence: 627-657 >POLY_AHSimple_repeat
Seq primer: M13 REVERSE.

FEATURES

source
1. .666
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajw-j-12-0-UI"
/clone_lib="UI-E-EJ1"
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ1 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6,791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTV73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (GT)₁₈ tail. The sequence tags for this library are: fetal eyes, ABAATCAAGA; lens, CGATTACCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT 143 a 160 c 185 g 175 t 3 others
ORIGIN

Query Match 97.5%; Score 316.8; DB 14; Length 666;
Best Local Similarity 98.8%; Pred. No. 7.9e-42;
Matches 318; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GAGGCGGCGAGGGGCTCGAGATCTCTCGACAGCAAGCCGCTCTGCTGCGCGCG 63
DB 328 GAGGCGGCGAGGGGCTCGAGATCTCTCGACAGCAAGCCGCTCTGCTGCGCGCG 387
QY 64 TCTCAGGGGCTGCTTCTCTCGAAATTGACGAGGGGTCTTTGGGAGAGCTGGCTT 123
DB 388 TCTCAGGGGCTGCTTCTCTCGAAATTGACGAGGGGTCTTTGGGAGAGCTGGCTT 447
QY 124 GAGGCGCTCATCCAGGCGAGGTTCTCGTTAGCTCTGTCGACCCACCTGCGGCGCTG 183
DB 448 GAGGCGCTCATCCAGGCGAGGTTCTCGTTAGCTCTGTCGACCCACCTGCGGCGCTG 507
QY 184 GCGTGAATCAGAAATTTTCCAAAGAGTAGTCTTTTGGCAAAACTCTAC 243
DB 508 GCGTGAATCAGAAATTTTCCAAAGAGTAGTCTTTTGGCAAAACTCTAC 567
QY 244 TTAATCCAAATGGGTTTTTCTCTGACAGTAGTTTCCAAATGTAATTAATTAACTA 303

DB 568 TTAATCCAAATGGGTTTTTCTCTGACAGTAGTCTCCAAATGTAATTAATTAACTA 627
QY 304 AAGTAAAAAAAAAAAAAAAAA 325
DB 628 AAGTAAAAAAAAAAAAAAAAA 649

RESULT 9
AM337157/c 614 bp mRNA linear EST 31-JAN-2000
LOCUS
DEFINITION xw81f06.x1 NCI CGAP Pauli Homo sapiens cDNA IMAGE:2834435 3', similar to gb:J05392 SYNDSCAN-1 PRECURSOR (HUMAN), contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION AM337157
VERSION AM337157.1 GI:6833783
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
EMAIL: cgapsb@mail.nih.gov
LIFE Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLNI at: www.bio.litl.gov/bdrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 359.

FEATURES

source
1. .614
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2834435"
/clone_lib="NCI CGAP Pauli"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1,72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT 157 a 178 c 151 g 127 t 1 others
ORIGIN

Query Match 97.2%; Score 315.8; DB 10; Length 614;
Best Local Similarity 99.4%; Pred. No. 1.2e-41;
Matches 317; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGGCGGCGAGGGGCTCGAGATCTCTCGACAGCAAGCCGCTCTGCTGCGCGCG 63
DB 319 GAGGCGGCGAGGGGCTCGAGATCTCTCGACAGCAAGCCGCTCTGCTGCGCGCG 260
QY 64 TCTCAGGGGCTGCTTCTCTCGAAATTGACGAGGGGTCTTTGGGAGAGCTGGCTT 123
DB 259 TCTCAGGGGCTGCTTCTCTCGAAATTGACGAGGGGTCTTTGGGAGAGCTGGCTT 200
QY 124 GAGGCGCTCATCCAGGCGAGGTTCTCGTTAGCTCTGTCGACCCACCTGCGGCGCTG 183
DB 199 GAGGCGCTCATCCAGGCGAGGTTCTCGTTAGCTCTGTCGACCCACCTGCGGCGCTG 140
QY 184 GCGTGAATCAGAAATTTTCCAAAGAGTAGTCTTTTGGCAAAACTCTAC 243
DB 139 GCGTGAATCAGAAATTTTCCAAAGAGTAGTCTTTTGGCAAAACTCTAC 80
QY 244 TTAATCCAAATGGGTTTTTCTCTGACAGTAGTTTCCAAATGTAATTAATTAACTA 303
DB 79 TTAATCCAAATGGGTTTTTCTCTGACAGTAGTTTCCAAATGTAATTAATTAACTA 20

```

OY 304 AAGTAAAAAAAAAAAAA 322
Db 19 AAGTAAAAAAAAAAAAA 1

RESULT 10
Al872681/c 367 bp mRNA linear EST 07-MAR-2000
LOCUS wn08e02.x1 NCI CGAP Paul Homo sapiens cDNA IMAGE:244858 3'
DEFINITION similar to gb:J05392 SYNDICAN-1 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION Al872681
VERSION Al872681.1 GI:5546730
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 367)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2055 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 191.
Location/Qualifiers
source 1..367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:244858"
/clone_lib="NCI CGAP Paul"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 93 a 95 c 86 g 93 t
ORIGIN
Query Match 97.1%; Score 315.6; DB 9; Length 367;
Best Local Similarity 98.8%; Pred. No. 1.7e-41;
Matches 318; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 GAGGGCGGCGAGGCTCTGAGATCTCTCGACACACGCCCTCTGCGCGCG 63
Db 324 GAGGGCGGCGAGGCTCTGAGATCTCTCGACACACGCCCTCTGCGCGCG 265
OY 64 TCTCCAGGGGCTGCTCTCTGGAATTGACGAGGGTGTCTTGGGACAGCTGGCT 123
Db 264 TCTCCAGGGGCTGCTCTCTGGAATTGACGAGGGTGTCTTGGGACAGCTGGCT 205
OY 124 GAGCGCTCCATCCAGGCGGCTCTCTGTTAGCTCTGTGGCCCACTGGGCGCTG 183
Db 204 GAGCGCTCCATCCAGGCGGCTCTCTGTTAGCTCTGTGGCCCACTGGGCGCTG 145
OY 184 GCGCTGGAATCAGAAATTTTCCAAAGAGTATGCTTTTGGTTTGGCAAACTCTAC 243
Db 144 GCGCTGGAATCAGAAATTTTCCAAAGAGTATGCTTTTGGTTTGGCAAACTCTAC 85
OY 244 TTAATCCAAATGGGTTTCTCTGTACAGTAGATTTTCCAAATTAACCTTTAATTA 303
Db 84 TTAATCCAAATGGGTTTCTCTGTACAGTAGATTTTCCAAATTAACCTTTAATTA 25
OY 304 AAGTAAAAAAAAAAAAA 325

```

```

Db 24 AAGTAAAAAAAAAAAAA 3

RESULT 11
BM968687/c 590 bp mRNA linear EST 21-MAR-2002
LOCUS UI-CF-DUI-aak-p-01-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
DEFINITION UI-CF-DUI-aak-p-01-0-UI 3', mRNA sequence.
ACCESSION BM968687
VERSION BM968687.1 GI:19586274
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 590)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-39, >AT rich/Low_complexity
Seq primer: M13 FORWARD
POLYA=yes.
Location/Qualifiers
source 1..590
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aak-p-01-0-UI"
/clone_lib="UI-CF-DUI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site: 1: EcoR I; Site: 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GCGCTAGGC.
TAG_Lib=UI-CF-DUI
TAG_TISSUE=Lung Epithelial Cells tissue nos 359-368
TAG_SEQ=GCGTAGGC"

BASE COUNT 153 a 173 c 146 g 118 t
ORIGIN
Query Match 97.1%; Score 315.6; DB 14; Length 590;
Best Local Similarity 98.8%; Pred. No. 1.3e-41;
Matches 318; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 GAGGGCGGCGAGGCTCTGAGATCTCTCGACACACGCCCTCTGCGCGCG 63

```

```

Db      322 GAGGCGCGCAGGCGCTGAGATCTCTCTGACAGACACGCCCGTCTGCTGTGGCGCG 263
QY      64 TCTCCAGGGGCTGCTTCTCTCTGAAATTTGACGAGGGGTGTCTTGGGCGAGACTGGCTCT 123
Db      262 TCTCCAGGGGCTGCTTCTCTCTGAAATTTGACGAGGGGTGTCTTGGGCGAGACTGGCTCT 203
QY      124 GAGCGCTCCATCCAGAGCGAGGTCTCTGTTAGCTCTGTGAGCCCGACCGTGGGCGCTG 183
Db      202 GAGCGCTCCATCCAGAGCGAGGTCTCTGTTAGCTCTGTGAGCCCGACCGTGGGCGCTG 143
QY      184 GCGTGGATCAGAAATTTTCCAAAGAGTAGTGTCTTGGTCTTGGCAAACTCTAC 243
Db      142 GCGTGGATCAGAAATTTTCCAAAGAGTAGTGTCTTGGTCTTGGCAAACTCTAC 83
QY      244 TTAATCCAGTGGTCTTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 303
Db      82 TTAATCCAGTGGTCTTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 23
QY      304 AAGTAAAAAAAAAAAAAAAA 325
Db      22 AAGTAAAAAAAAAAAAAAAA 1

RESULT 12
LOCUS   A1224622 548 bp mRNA linear EST 21-DEC-1998
DEFINITION gw56n04.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:1998967 3' similar to gb:J05392 SYNDSCAN-1 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION A1224622
VERSION   A1224622.1 GI:3807335
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
unknown library type
Insert Length: 265 Std. Error: 0.00
Seq primer: -40UP from Gldco
High quality sequence stop: 352.
Location/Qualifiers
FEATURES
source 1..548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1998967"
/clone_lib="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with signed ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NciI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 134 a 158 c 141 g 109 t 6 others
ORIGIN
Query Match 96.7%; Score 314.4; DB 9; Length 548;
Best Local Similarity 99.7%; Pred. No. 2.1e-41;
Matches 315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GAGGCGCGCAGGGGCTGAGATCTCTCTGACACACGCCCGTCTGCTGTGGCGCG 63
Db      316 GAGGCGCGCAGGGGCTGAGATCTCTCTGACACACGCCCGTCTGCTGTGGCGCG 257
QY      64 TCTCCAGGGGCTGCTTCTCTCTGAAATTTGACGAGGGGTGTCTTGGGCGAGACTGGCTCT 123
Db      256 TCTCCAGGGGCTGCTTCTCTCTGAAATTTGACGAGGGGTGTCTTGGGCGAGACTGGCTCT 197

```

```

QY      124 GAGCGCTCCATCCAGAGCGAGGTCTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 183
Db      196 GAGCGCTCCATCCAGAGCGAGGTCTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 137
QY      184 GCGTGGATCAGAAATTTTCCAAAGAGTAGTGTCTTGGTCTTGGCAAACTCTAC 243
Db      136 GCGTGGATCAGAAATTTTCCAAAGAGTAGTGTCTTGGTCTTGGCAAACTCTAC 77
QY      244 TTAATCCAGTGGTCTTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 303
Db      76 TTAATCCAGTGGTCTTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 17
QY      304 AAGTAAAAAAAAAAAA 319
Db      16 AAGTAAAAAAAAAAAA 1

RESULT 13
LOCUS   AM192078 519 bp mRNA linear EST 29-NOV-1999
DEFINITION x179E03.x1 NCI CGAP Pauli Homo sapiens cDNA clone IMAGE:2680925 3' similar to gb:J05392 SYNDSCAN-1 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION AM192078
VERSION   AM192078.1 GI:6470777
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at: www.bio.liml.gov/bbtp/image/image.html
Seq primer: -40UP from Gldco
High quality sequence stop: 321.
Location/Qualifiers
FEATURES
source 1..519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2680925"
/clone_lib="NCI CGAP Pauli"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NciI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

BASE COUNT 127 a 153 c 131 g 108 t
ORIGIN
Query Match 96.6%; Score 314; DB 10; Length 519;
Best Local Similarity 98.4%; Pred. No. 2.6e-41;
Matches 317; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      4 GAGGCGCGCAGGGGCTGAGATCTCTCTGACACACGCCCGTCTGCTGTGGCGCG 63
Db      322 GAGGCGCGCAGGGGCTGAGATCTCTCTGACACACGCCCGTCTGCTGTGGCGCG 263
QY      64 TCTCCAGGGGCTGCTTCTCTCTGAAATTTGACGAGGGGTGTCTTGGGCGAGACTGGCTCT 123
Db      262 TCTCCAGGGGCTGCTTCTCTCTGAAATTTGACGAGGGGTGTCTTGGGCGAGACTGGCTCT 203
QY      124 GAGCGCTCCATCCAGAGCGAGGTCTCTCTGTAAGTAGATTTTCCAAATGTAATACTTAATA 183

```

Db 202 GAGCGCTCCATCCAGAGCCAGGTTCTCCGTAGCTCTGTG96CCCCCACCCTG36CCCTG 143

Qy 184 GGCTGGAATCAGAGATATTTTCCAAAGAGTGA TAGCTTTTGGCTTTGGCAAACTTAC 243

Db 142 GGTGGAAATCAGAGATATTTTCCAAAGAGTGA TAGCTTTTGGCTTTGGCAAACTTAC 83

Qy 244 TTAATCCAAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATTAACCTTAATA 303

Db 82 TTAATCCAAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATTAACCTTAATA 23

Qy 304 AAGTAAAAAATTTTAAAAA 325

Db 22 AAGTAAAAAATTTTAAAAA 1

RESULT 14

LOCUS AM62594 461 bp mRNA linear EST 06-APR-2000

DEFINITION h13a10.x1 NCI CGAP Col4 Homo sapiens cDNA clone IMAGE:2974074 3' similar to gb:J05392 SYNDACAN-1 PRECURSOR (HUMAN); mRNA sequence.

ACCESSION AM62594

VERSION AM62594.1 GI:7455133

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: image.llnl.gov/image/html/resources.shtml

FEATURES

source

1. 461

Location/Qualifiers

/db_xref="taxon:9606"

/db_xref="IMAGE:2974074"

/clone_lib="NCI-CGAP Col4"

/tissue_type="moderately-differentiated adenocarcinoma"

/lab_host="DH10B"

/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 Kb. Life Technologies catalog #: 11531-019"

BASE COUNT 112 a 123 c 118 g 108 t

ORIGIN

Query Match 96.5%; Score 313.6; DB 10; Length 461; Best Local Similarity 98.8%; Pred. No. 3.2e-41; Matches 316; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 AAGGCGGAGGAGGCTGAGATCTCTCGACAGCCAGCCGCTGCTGGGCGCGT 64

Db 320 AAGGCGGAGGAGGCTGAGATCTCTCGACAGCCAGCCGCTGCTGGGCGCGT 26;

Qy 65 CTCGAGGAGGCTCTCTCTCTGGAATTCAGAGGAGTCTTGGCAGAGCTGGCTTG 124

Db 260 CTCGAGGAGGCTCTCTCTCTGGAATTCAGAGGAGTCTTGGCAGAGCTGGCTTG 201

Qy 125 AACGCGTCCATCCAGAGCCAGGTTCTCCGTAGCTCTGTG96CCCCCACCCTG36CCCTG 184

Db 200 AACGCGTCCATCCAGAGCCAGGTTCTCCGTAGCTCTGTG96CCCCCACCCTG36CCCTG 141

Qy 185 GGTGGAATCAGAGATATTTTCCAAAGAGTGA TAGCTTTTGGCTTTGGCAAACTTAC 244

Db 140 GGTGGAATCAGAGATATTTTCCAAAGAGTGA TAGCTTTTGGCTTTGGCAAACTTAC 81

Qy 245 TTAATCCAAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATTAACCTTAATA 304

Db 80 TTAATCCAAATGGGTTTCTCTGTACAGTAGATTTTCCAAATGTAATTAACCTTAATA 21

Qy 305 AAGTAAAAAATTTTAAAAA 324

Db 20 AAGTAAAAAATTTTAAAAA 1

RESULT 15

LOCUS A1691025 636 bp mRNA linear EST 14-DEC-1999

DEFINITION A1691025 tgl3e11.x1 NCI CGAP Utr3 Homo sapiens cDNA clone IMAGE:2208716 3' similar to gb:J05392 SYNDACAN-1 PRECURSOR (HUMAN); contains element PIR5 repetitive element ; mRNA sequence.

ACCESSION A1691025

VERSION A1691025.1 GI:4902327

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: www.bio.llnl.gov/bbtp/image/image.html
Insert Length: 1849 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 337.

FEATURES

source

1. 636

Location/Qualifiers

/db_xref="taxon:9606"

/db_xref="IMAGE:2208716"

/clone_lib="NCI CGAP Utr3"

/tissue_type="poorly-differentiated endometrial adenocarcinoma, 2 pooled tumors"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.45 Kb. Life Technologies catalog #: 11541-018"

BASE COUNT 157 a 173 c 152 g 154 t

ORIGIN

Query Match 96.1%; Score 312.4; DB 9; Length 636; Best Local Similarity 98.1%; Pred. No. 4.1e-41; Matches 316; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 4 GAGGCGGAGGAGGCTGAGATCTCTCGACAGCCAGCCGCTGCTGGGCGCGT 63

Db 345 GAGGCGGAGGAGGCTGAGATCTCTCGACAGCCAGCCGCTGCTGGGCGCGT 286

Qy 64 TCTCCAGGAGGCTCTCTCTCTGGAATTCAGAGGAGTCTTGGGCAAGAGCTGGCTCT 123

